

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:21:40 ; Search time 127 Seconds

(without alignments)
369.954 Million cell updates/sec

Title: US-09-917-384-4

Perfect score: 2249
Sequence: 1 ATHVDNPPYAGATFFVNPYWA.....ACQWPPAQFDQIVANAPAV 423

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

1: A_Geneseq_032802.*
2: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1999.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1023	45.5	1010	19	AAW34989	Teredinbacter end
2	999	44.4	423	17	AAR90716	Thermostable cellu
3	999	44.4	596	17	AAR90715	Thermostable cellu
4	569	25.3	395	21	AAV84797	Amino acid sequenc
5	562.5	25.0	476	19	AAW44853	Humicola insolens
6	562.5	25.0	476	19	AAW44827	H. insolens celluli
7	558.5	24.8	473	20	AAV01077	H. insolens Cel6a
8	546	24.3	471	6	AAV50308	Cellobiohydrolase
9	538	23.9	471	16	AAR77262	T. longibrachiatum
10	538	23.9	471	17	AAW02025	Trichoderma cellulo
11	534	23.7	457	18	AAW25789	Acromonium cellulo

12	449.5	20.0	446	20 AAV01076	H. insolens Cel6B
13	439.5	19.5	449	19 AAW56739	Orpinomyces celluli
14	431.5	19.2	432	21 AAB18414	Amino acid sequenc
15	411	18.3	459	19 AAW56738	Orpinomyces celluli
16	244.5	10.9	449	18 AAW18210	Cellobionces flm1
17	231	10.3	490	20 AAW95602	Saccharothrix aust
18	181.5	8.1	321	22 AAB70839	S. halstedii cellu
19	120.5	5.4	1790	22 AAB71739	Drosophila melanog
20	108.5	4.8	1095	10 AAP97052	Sequence encoded b
21	107	4.8	639	22 AAE12935	Dactylium dendroid
22	107	4.8	639	22 AAE12935	Dactylium dendroid
23	107	4.8	639	22 AAE12939	Dactylium dendroid
24	107	4.8	639	22 AAE12941	Dactylium dendroid
25	107	4.8	680	22 AAC63582	Amino acid sequenc
26	107	4.8	681	22 AAC63583	Synthetic amino ac
27	106	4.7	639	22 AAE12964	Dactylium dendroid
28	105.5	4.7	639	22 AAE12962	Dactylium dendroid
29	105	4.7	632	16 AAR77558	p45 metallopeptase
30	105	4.7	639	22 AAE12953	Dactylium dendroid
31	104	4.6	7068	22 AAE12960	Dactylium dendroid
32	104	4.6	639	22 AAE12966	Dactylium dendroid
33	104	4.6	9477	22 AAE10144	Dactylium dendroid
34	103.5	4.6	1472	22 ABB59423	Dactylium dendroid
35	103	4.6	639	22 AAE12957	Dactylium dendroid
36	103	4.6	639	22 AAE12940	Dactylium dendroid
37	103	4.6	639	22 AAE12946	Dactylium dendroid
38	103	4.6	639	22 AAE12966	Dactylium dendroid
39	102	4.5	639	22 AAE12934	Dactylium dendroid
40	101.5	4.5	639	22 AAE12928	Dactylium dendroid
41	101	4.5	384	21 AAC64647	Arabidopsis thalia
42	101	4.5	421	21 AAC64647	Arabidopsis thalia
43	101	4.5	502	21 AAC64647	Arabidopsis thalia
44	100	4.4	1888	22 AAU31049	Novel human secret
45	99.5	4.4	388	17 AAW05845	Fusarium oxysporum

ALIGNMENTS

RESULT 1	AAW34989 standard; Protein; 1010 AA.
AAW34989;	
21-MAY-1998 (first entry)	
Teredinbacter endoglucanase:	
Endoglucanase; cellulase; carboxymethylcellulose; cellulose;	
biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;	
thermostable enzyme; thermophilic; glycosidase.	
Teredinbacter sp. (Clone 426P1).	
MO9744361-A1.	
27-NOV-1997.	
22-MAY-1997;	97WO-US08793.
22-MAY-1996;	96US-0651572.
(RECO-) RECOMBINANT BIOCATALYSIS INC.	
Lam DE, Mathur EJ;	
WPI: 1998-018435/02.	
N-PSDB; AAT94197.	
Endoglucanase(s), preferably from archaeal bacterium, AEPit 1a -	
useful to degrade carboxymethylcellulose and hydrolyse of	
beta-1,4-glycosidic bonds in cellulose	

BEST AVAILABLE COPY

```
XX Claim 1; Fig 1E; 16app; English.
PS
XX
CC This protein comprises an endoglucanase of Terebinthabacter (clone
CC 426P1) that is capable of degrading carboxymethylcellulose and of
CC hydrolysing the beta-1,4-glycosidic bonds in cellulose. It has
CC homology to an endoglucanase of archaeobacterium ABE111 (see
CC AAW34985). It can be produced from native cells or from recombinant
CC host cells, especially prokaryotic host cells transformed with a
CC plasmid or virus-derived vector including the endoglucanase DNA
CC (see AAT94197). 24 Endoglucanases (see AAW34986-W35008) are claimed.
CC They can be used to degrade cellulose for the conversion of plant
CC biomass into fuels and chemicals, for use in detergents, textiles,
CC animal feed, waste treatment, and in the fruit juice and brewing
CC industries for the clarification and extraction of juices.
XX
SQ Sequence 1010 AA;

Query Match 45.5%; Score 1023; DB 19; Length 1010;
Best Local Similarity 48.5%; Pred. No. 2.3e-78;
Matches 210; Conservative 68; Mismatches 121; Indels 34; Gaps 13;

OY 4 VDNPYAGA-TFVNPFYMAQVQSEANQNTATLAAKMRVYSTSTAVWMDRIAINGVNG 62
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 598 vdnpyagaakqkylmpms-----asaanepps-----vianeptfymndrigalegpad 647
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 63 GGLGTLTYLDALISQOOGTTPVEIEIYIDLPGRCALIASNGELPATYAGLOTYETQYID 122
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 648 gngldthdhealaq---gadlfmfvvdlpnrdcaalaangelrlsedgfnlyksdyla 703
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 123 PFIASISNPKYSLSRLVITIEPDSLPMNATNMKSIOAC--ATAVPYEGGIEVALTRKLMH 180
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 704 pfielsldpaysgilaavaievdslpnlvtntsepdceqanpggyrdgltlmalteigkrl 763
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 181 PNVTYTMDAASHGWLGPNN-ASGYVOEVQKVLNASIGVNGIDGVTNTANTYTPKEPPM 239
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 764 pnyvyvdlahgswlqgndfnagynlllyevaanlgsjnpdlagfysnaanytlpreeql 823
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 240 -TATQOVGGOPVESANFYQWNPDIIDEADVAVULYSRLVAAAGPPSSIGMLIDTLRNGWGP 298
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 824 pdaalqvgqpyrsadfyewmsylaelkpfvlrwsamlskmpsslgmlidtlarnwgwp 883
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 299 NEPTGPSTADTFTVFNOSKIDLRHRLGCMNGAGLGPQASPTDEPNMHLDAVYVMI 358
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 884 eptgsgstsmnlntfvesrlrtrehrnwmcpqg-9vaylptaa---pspgldayvww 938
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 359 KPGESEDGTSAS---DPT-TGKKSDDPMCDP---TYSYGVLTNALPNSDIAGOWPFA 410
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 939 kpgesdgysdpndfeldpndpkqhdpmcdpfasnsnsaayg--lgampnaphagrtwfrpe 996
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 411 QPDOLVANAARPAV 423
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 997 atqlllenayppl 1009
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 2
AAR90716
ID AAR90716 standard; Protein; 423 AA.
AC AAR90716;
DT 06-APR-1996 (first entry).
DE Thermostable cellulase-E3 catalytic domain.
KW Cellulase-E3; Thermomonospora fusca; thermostable; papain;
KW catalytic domain; Escherichia coli; cloning; Streptomyces lividans;
KW cellulose; chitosan; protease resistance; synergism;
KW cellobiohydrolase; beta-glucosidase; saccharification;
KW surfactant additive; paper recycling; delinking; paper refining.
OS Thermomonospora fusca strain YX36.
```

```
XX
XX W09600281-A1.
XX
XX 04-JAN-1996.
XX
XX 23-JUN-1995; 95WO-US09069.
XX
XX 24-JUN-1994; 94US-0265429.
XX
XX (CORR ) CORNELL RES FOUND INC.
XX Walker LP, Wilson DB, Zhang S;
XX WPI: 1996-068865/07.
XX N-PSDB; AAT15596.
XX
XX Cellulase from Thermomonospora fusca - is thermostable and is useful
XX in a variety of industrial applications e.g. clarification of fruit
XX juices, fabric softening, etc
XX
XX Claim 1; Page 38-40; 53pp; English.
XX
XX The sequence corresponds to a catalytically active domain of
XX thermostable cellulase-E3 (EC-3.2.1.4) from Thermomonospora
XX fusca, and is isolated by papain cleavage of the full-length
XX protein (AAR90715). The binding domain and linker region have
XX been removed from the full-length protein. The protein may be
XX produced in recombinant form in Escherichia coli or Streptomyces
XX lividans. The domain retains full activity, but shows reduced
XX cellulose binding. The catalytic domain, or the corresponding
XX full-length cellulase, hydrolyses cellulose or chitosan at pH
XX 5-11 and 40-70 deg C, and has significant activity at over 60
XX deg C. E3 has higher stability to proteolysis in culture
XX supernatant than T. fusca cellulases-E2 and -E5, and shows strong
XX synergistic activity when combined with other cellulases,
XX cellobiohydrolase and optionally beta-glucosidase. The cellulase
XX and mixtures may be used e.g. in cellulose saccharification for
XX ethanol production, fruit juice clarification, as a surfactant
XX additive, or in delinking or refining of recycled paper.
XX
XX Sequence 423 AA;

Query Match 44.4%; Score 999; DB 17; Length 423;
Best Local Similarity 48.3%; Pred. No. 7.2e-77;
Matches 209; Conservative 55; Mismatches 137; Indels 32; Gaps 9;

OY 4 VDNPYAGATFVNPFYMAQVQSEANQNTATLAAKMRVYSTSTAVWMDRIAINGVNG 59
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 6 vdnpyagaaklynpwksaaaepps-----avaanstavldrtigalegndsp 55
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 60 VNGGRLTTYLDALISQOOGTTPVEIEIYIDLPGRCALIASNGELPATYAGLOTYETQ 119
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 56 tngsmglrldheav-rsgsgdplrlqvlylnlprdrcaalaangelgde--ldrykee 112
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 120 YIDPFIASISLN-PKYSLSRLVITIEPDSLPMNATNMKSIOACATNPY-----YEGGIEY 172
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 113 yidpfiadlmvdiadynllivallleldsipnlvtvngngytelecaymkqngyynygy 172
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 173 ALTKLAIIPNVTYMDAASHGWLGPNNASGYVOEVQKVLNS-IGVNGIDGVTNTATY 231
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 173 alrklgeipnyynidaahgwlqwdnsfgyvdllyeaansgstvdvvhfistatny 232
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 232 TPKEPEMTATQOVGGOPVESANFYQWNPDIIDEADVAVULYSRLVAAAGPPSSIGMLIDTL 291
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 233 satpeyldvngvtngqlrsgskwdwngydelstfvgqlrallakgrfsgldlids 292
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 292 RKGWGPNNPTGPSTADTFTVFNOSKIDLRHRLGCMNGAGLGPQASPTDEPNMHL 351
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 293 rkgwgpnnptgppsstdlntvdesdrtrllhpgnwmcqaqaglgreptvpa---pg 348
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 352 LDAYWIKRPGESEDGTSASDPTTGKKSDDPMCDPRTYTSYGVLTN---ALPSPSIAGOWF 408
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : :
```


PT Mutant Chrysosporium strain comprising nucleic acid encoding proteins,
 particularly industrially important enzymes such as cellulases,
 xylanases, pectinases, lipases and proteases -

PS Disclosure; Page 43-46; 92pp; English.

XX The present sequence represents a Chrysosporium C1 endoglucanase. The
 CC polynucleotide sequence comprises a promoter sequence, and can be
 CC linked to a nucleic acid sequence of the invention to control
 CC expression. The specification describes a mutant Chrysosporium
 CC strain which comprises a nucleic acid sequence encoding a polypeptide
 CC of interest, where the nucleic acid sequence is linked to an
 CC expression-regulating region and a secretion signal sequence.
 CC The mutant strain is useful for producing heterologous proteins
 CC and polypeptides of interest, e.g. industrially useful enzymes such
 CC as cellulases, xylanases, pectinases, lipases and proteases.

CC Sequence 395 AA;

Query Match 25.3%; Score 569; DB 21; Length 395;
 Best Local Similarity 35.5%; Pred. No. 3.7e-40;
 Matches 150; Conservative 54; Mismatches 148; Indels 70; Gaps 13;

OY 1 ATHVNDNPRYAGATFVNPRYAOEVSAAQNTNATLAAKRWVSTYSTAVMMRIAINGV 60

DB 43 stnrfqgy---tlhpnrfyaeeaaaeaisdsalaekarkvadvgfllwltlenlg-- 97

OY 61 NGGPGILTYTLDALSOQOGTTPREIVETIVYDUPGRDCAALASNGELPATNAGLQTEYTOY 120

DB 98 -----rlpeal--edpcenivglviydliprdcaakaangel--kvgeldrykley 145

OY 121 IDPIASILSNPKYSSRIYTTIEPDSLPAVNTNMSIQACATVPEEGIEYALTKLHAI 180

DB 146 ldkiaelk--ahntafalviepdsipolvnsdlqtcqgsaagyregvayalkqln-1 202

OY 181 PNVTIYMDAASHGMLGMPNNAAGVQEOVGKINASTIGVNGIDGFPVTNTNANPRKPEPMT 240

DB 203 pnvmyidagbhgwlgwvndnlprgaqelasyksaagspqvvgistnvaqw-----n 254

OY 241 ATQVGGQPVESANFYQWNPIDIDADVADLYSRLVAGFPSSIGMLIDTLRNGMGPNNE 300

DB 255 awdggpge-fsdasdaqykcqnekylntfgaelksagmhp--aldtgnrytg--- 308

OY 301 PRGSPATDVNFVNSKIDLEQHGKMGONGAGIGOPPOASPTDFPVAHLDAYWITRP 360

DB 309 -----lrdewgdwcunvgagfgvyrpantgd--eladafwvxp 345

OY 361 PGESDGTSAADPTTGKKSDDPTTYTTSYGVLTNNALNSPIAGOWPFAOPDOLVANAR 420

DB 346 ggesdgtdsds---aarydsfcgk-----pdafkpspeagtwngayfemllknan 392

OY 421 PA 422

DB 393 ps 394

RESULT 5

ID AAM44853 standard; Protein; 476 AA.

AC AAM44853;

XX 31-JUL-1998 (first entry)

DE Humicola insolens cellulase NCE2 protein.

XX Humicola insolens; NCE1; NCE2; NCE4; cellulase; expression vector;

KM promoter; signal sequence; terminator; amylase; lipase; protease;

XX Humicola insolens.

FN Key Location/Qualifiers
 FT Peptide 1..23
 FT /label= signal
 FT Protein 23..476
 FT /label= Cellulase_NCE2

PN W09803667-A1.

PD 29-JAN-1998.

XX 24-JUL-1997; 97WO-JP02560.

XX 24-JUL-1996; 96JP-0195070.

XX (MEIJI) MEIJI SEIKA KAISHA LTD.

PA Aoyagi K, Hamaya T, Koga J, Kono T, Moriya T, Murakami T;

PI Muroshima K, Sumida N, Watanabe M;

XX WPI; 1998-120786/11.

DR N-PSDB; AAV19377.

PT Mass production of proteins and peptides in Humicola species - using

PT expression vector containing the promoter, signal sequence and/or

PT terminator from the Humicola insolens NCE1 or NCE2 gene

PS Claim 8; Page 40-45; 63pp; Japanese.

XX The present sequence represents the Humicola insolens cellulase NCE2
 CC protein from the present invention. The present invention describes a
 CC method for the mass production of proteins and peptides in Humicola
 CC species, especially in Humicola insolens, using an expression vector
 CC which comprises the promoter, signal sequence and/or terminator
 CC regulatory sequences from the NCE1 or NCE2 gene of H. insolens. These
 CC are available in the plasmids pM3-1 (Escherichia coli JM109/pM3-1,
 CC FERM BP-5971) (for NCE1) and pM14-1 (E. coli JM109/pM14-1, FERM BP-5972)
 CC (for NCE2). The vector also contains a marker gene such as an antibiotic
 CC resistance gene (e.g. the tetracycline resistance gene from Streptomyces
 CC limofaciens). Proteins which can be expressed using this system include
 CC cellulase, amylase, lipase, protease, phytase and other enzymes.
 CC Specific expression vectors of the invention are pMKD01 (for Humicola
 CC NCE3 cellulase gene), pEGD01 (for Humicola NCE4 cellulase gene) and
 CC pHEM02 (for Humicola NCE4 cellulase gene). The expression system allows
 CC the efficient production of proteins and peptides in a Humicola host.
 CC Using the expression system high amounts of protein (>4.5 g/l) can be
 CC obtained.

SO Sequence 476 AA;

Query Match 25.0%; Score 562.5; DB 19; Length 476;
 Best Local Similarity 33.7%; Pred. No. 1.7e-39;
 Matches 147; Conservative 54; Mismatches 146; Indels 89; Gaps 15;

OY 1 ATHVNDNPRYAGATFVNPRYAOEVSAAQNTNATLAAKRWVSTYSTAVMMRIAINGV 59

DB 113 aeyngnprlegvqlwvannnyrsewhltalpqldpalaasaavaepsfqwidrvtyd- 171

OY 60 VNGGPGILTYTLDALSO-----QOGTTPRE-VIEIYIYDUPGRDCAALASNGELPATNAGL 113

DB 172 -----tlivetsetraangagannpyaaqiyvdlpdrdcaaaasngewalamnga 223

OY 114 QTEFYQYIDPIASILSNPKYSSRIYTTIEPDSLPAVNTNMSIQACATVPEEGIEYALTKLHAI 173

DB 224 nnyk-gylirirellis--fsdvrtllviepdsalnmvnmvaksagsaasyrrelltlya 280

OY 174 LTRKHAIPNVITYMDAASHGMLGMPNNAAGVQEOVGKINASTIGVNGIDGFPVTNTNANPRKPEPMT 232

DB 281 lqyld-lphvanymdagbhgwlgwvndnlprgaqelasyksaagspqvvgistnvaqw----- 339

OY 233 -----PLKEPMTATQVGGQPVESANFYQWNPIDIDADVADLYSRLVAGFPSSIGMLIDTLRNGMGPNNE 285

DB 340 waispppytsp-----npydekhyieafrrllleargfpaq-- 376


```
OY      286 MIDILRLNMGSPNEPTGPSTADVTWTFVNSQKIDLRHGRGLMCNONGACLGOPQPASPT   345
       ::|||::|||||
Db      377 flvdvgdgsr-----kqptg-----qkgwghncnaigtcfgmprant-    413
OY      346 DEFPNAHLDAIVIKRPGESDGTSAASDPYTGGKKSDPMCDPFTYTSYGVLTNALPNPSIAG   405
       ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      414 -_ghvyvddftvvvkpgsgedgt----dttaarydyhc-----ledalkpapaaq     458
OY      406 GMEPAOFODLVANARP        421
       ||| | |:|::| | | |
Db      459 qwfgyfeqlrltnarp         474

RESULT          6
ID AAM44827 standard; Protein: 476 AA..
XX AAM44827;
XX AAW44827;
Dt      07-JUL-1998 (first entry)
DE H. insolens cellulase NC2 protein sequence.
XX Cellulase; primer: PCR; amplification; probe; hybridisation; hydrolysis;
KM cellulose; sugar; enzyme.
OS Humicola insolens.
XX
FH Key Location/Qualifiers
ET Peptide 1..23
ET /note= "signal peptide"
FT Protein 24..476
FT /note= "NC2 cellulase mature protein"
PX JP08126492-A.
PD 21-MAY-1996.
FX 02-NOV-1994; 94JP-0269293.
PR 02-NOV-1994; 94JP-0269293.
XX (MEIJ ) MEIJI SEIKA KAISHA LTD.
DR WPI: 1998-056048/06.
DX N-PADB: AAV19281.
XX Novel cellulase component NCE2 - Isolated from Humicola insolens and
Pt used in detergents, as a feed additive etc.
Pr Discloure: Page 10-12; 15pp; Japanese.
PS This is the amino acid sequence of the NC2 cellulase from Humicola
CC Insolens. The protein was isolated from cultured cells and the
CC N-terminal peptide sequence determined. The protein was also
CC cleaved with V8 protein to generate several peptide fragments. These
CC were also sequenced. Degenerate primers were designed based on the
CC sequences of the protease fragaments (see AAV19282-V19283) and used to
CC amplify a fragment of the gene for use as a probe to screen a genomic
CC library. The remainder of the sequence was obtained and determined using
CC primers AAV19284-V19299. The cellulase can be used to hydrolyse
CC cellulose into sugar, and as an enzyme in industry, as major component
CC in a wide variety of products such as detergent, fibre-processing agent,
CC feed additive, or digestive agent.
```

Query Match	25.0%;	Score 562.5;	DB 19;	Length 476;
Best Local Similarity	33.7%;	Pred. NO. 1.7e-39;		
Matches 147; Conservative	54;	Mismatches 146;	Indels 89;	Gaps 15

```

Oy      1 ATHNDNPAAGTFFPVNNYMAOEVOSEAAQ--TNATLAAKRRVSTYSTVAWMDRIJAAIG   59
Db      113 asyngnpfegvqldvaannyysenhtlaipdltcpalraasaavaevpsfqlndrntvd-  171
Oy      60 VNGRGGLTTYDAALSQ-----QQGTPPE-VIEIVTYDLPRGDCAALASNGELPATACL  113
Db      172 -----tliveteliraangagangnpyaedlvyvdlprdtcaaaasngewalaanga  223
Oy      114 QTEYOXYIDPIASTLSNPRTXSRLRYVIIEPDSLPAVNFMMSIQACATVPVYEQSIGEA  173
Db      224 nnyk-gyanlirellis--fsdvrtllivapdsldamvtlmnnvakcsagaascylreltya  280
Oy      174 LTKLHAIPNVYIYDAAHSGMLGNPNNASGYVOEVOKLVLNASIGVGIDGFVTNTANTY-  232
Db      281 lkqdl-lphvamyndaghagwlgypnanlqpaaelfkalyedagkravzgatunanya  339
Oy      233 -----PLKEPFMTAQOQGGOGVESANTRYQNRPDIDEDADVLDXSLVLAAGFPSSSIG  285
Db      340 wslsspppytsp-----npydekhyleatrpjlleargfpdq--  376
Oy      286 MLIDTLRNGWGSPNEPTGPSTATDVNTFVNOQSRTIDLRQHKGLCNCNONGALCOPPOASPT  345
Db      377 flvdgsgtsq----kqrpgs-----qkwghbcmnlglvgfgmfrplac-  413
Oy      346 DFNNAHLDAVYWIKRPESDSGTSAADPTTGKKSDDPCDEFTYTTSIGVLTNALPNSPIAG  405
Db      414 --gbqyvdaftwvkprgsegcdgt---dtlaarydyhcg-----ledalkpapeeg  458
Oy      406 QMPPAOFDOULVANARP 421
Db      459 qwfqayfeqllrnamp 474

```

RESULT 7

ID AAY01077 standard; Protein; 473 AA

AC AAY01077

DT 08-JUN-1999 (first entry)

H. insolens Cel6A fungal cellulase protien sequence

KW Cel6B; Cel6A; fungal cellulase; cleaning composition; conditioner;

KW fabric softener; Humicola-like cellulase; glycosyl hydrolase family 6;

XX
XX
XX

[illegible]XX
14 JAN 1966

XX
XX
C
THH
T
C
C
C
C
C
C
C
C
C

[illegible][illegible][illegible]

XX
XX
XXXX-XXXXXXX

DR N-PSDB; AAX27948, AAX27958.

PT Cleaning composition containing Humicola endo-beta-1,4-glucanase

PT	conditioners
xy	

Example 3; Page 243-245; 271pp; English

CC This sequence is the Humicola insolens Cel6A fungal cellulase.


```

XX AAR77262;
AC 13-DEC-1995 (first entry)
DT T. longibrachiatum exo-cellulohydrolase CBHII.
DE Cellulase; cellulose; signal; catalytic core; cellulase binding;
KW linker.
XX Trichoderma longibrachiatum.
OS
XX Key Location/Qualifiers
FH Peptide 1..24
FT /label= signal
FT /note= "seq id no 28"
FT Domain 25..46
FT /label= cellulose binding domain
FT /note= "seq id no. 4"
FT Region 47..89
FT /label= linker
FT /note= "seq id no 20"
FT Domain 90..471
FT /label= catalytic core
FT /note= "seq id no 12"
XX WO9516782-A.
XX 22-JUN-1995.
XX 19-DEC-1994; 94WO-US14163.
XX 17-DEC-1993; 93US-0169948.
XX (GEMV ) GENENCOR INT INC.
XX Clarkston KA, Collier KD, Fowler T, Larenas E, Ward M;
XX N-PSDB; AAO91282.
XX WPI; 1995-231574/30.
XX DR N-PSDB; AAO91282.
XX Pure, truncated fungal cellulase protein from Trichoderma - useful to
XX reduce or eliminate dye, colourant or pigment back-staining or
XX redeposition in stone-washing or bio-polishing
XX Claims 12, 39, 40, 41; Figure 2; 105pp; English.
XX
XX Figure 2 depicts the genomic DNA and AA sequence of CBHII derived
XX from T. longibrachiatum. A truncated fungal cellulase with the AA
XX sequence in SEQ ID no. 4 is claimed. Also claimed are DNA gene
XX fragments encoding sequences: SEQ ID 12; SEQ ID 12 and 20; SEQ
XX ID 12, 20 and 16. Genes for CBHII have been isolated
XX from T. longibrachiatum and the protein domain structure has been
XX confirmed (Shoemaker, S. et al., 1983, Bio/Technology 1, 691-696;
XX Teeri, T. et al., 1983, Bio/Technology 1, 696-699 and Teeri, T. et al.,
XX 1987, Gene, 51, 43-52).
XX
XX Sequence 471 AA:

```

```

Query Match 23.9%; Score 538; DB 16; Length 471;
Best Local Similarity 33.3%; Pred. No. 2..1e-37;
Matches 144; Conservative 51; Mismatches 155; Indels 82; Gaps 13;

```

```

OY 1 ATRHVDNRYGATFFVNPYNAQEQSEANOTNATLAARKRVSTYTAVMMD-----52
DB 110 ATRHVDNRYGATFFVNPYNAQEQSEANOTNATLAARKRVSTYTAVMMD-----52
OY 53 --RIALINGVNGGPGGLTWTIDALSOQOGTTPEVIEIYIDJGRCACALASGEIPAPA 110
DB 170 eqtladitranngg--nyag-----qfvdldpdrcaalaasngeslad 213
OY 111 AGLOTYETOYIDPIASILSNPKYSRLIVTITPEDSLPAVNTMSIQACATAVPEYEGGI 170

```

```

DB 214 ggvakyk-nyidtlrqlv--veysdlrtllvlepdslanlvtnlgtprkcanagsaylecl 270
OY 171 EYALTKLHALIPNYIYIMDAHSGMLGMPNNAASGYVEQVKVLNASTGCVNIDCFVNTAN 230
DB 271 nyavtqln-1pnyamyldqghagwlgspanqdpaaqlfanvyxnaasprallrglatnvan 329
OY 231 YTPLEKPEFMATQOVGQEPESANFYQWNPDIADYAVDLYSRLVAAGFPSSIGMLIDT 290
DB 330 yn-----gnwltspspytqgnaynekylinaigpllanhgv-snaftldq 375
OY 291 LRNGMGPNPEPTGPTATVDVNTFVNOSKIDLRQHRGLMCNQCAGLQPPQASPTDFPNA 350
DB 376 grrg-----kqptg-----ggqwgdcwcnvlgfgrprsanltdg---s 410
OY 351 HLDAYVWIKRPGSGDSGTSAADPTTGKSDPMCDPTTYSYGLTNALPNSPIAQMFPRA 410
DB 411 llsfvmvprpggcdgldssap-----fidsnc-----alpdalqpapagawfqa 457
OY 411 QPDQLVAMARPA 422
DB 458 yfvqlltnanps 469

```

```

RESULT 10
AAW02025
ID AAW02025 standard; Protein; 471 AA.
XX
XX AAW02025;
XX 28-OCT-1996 (first entry)
XX Trichoderma cellulohydrolase II.
XX Cellulohydrolase II; CBHII; cellulase; cellulose; dentin;
XX stonewashing; dye redeposition; backstaining.
XX Trichoderma longibrachiatum.
XX
XX Key Location/Qualifiers
FH Peptide 1..24
FT /label= sig_peptide
FT Protein 25..471
FT /label= Mat_protein
FT Domain 25..63
FT /label= Cellulose_binding_domain
FT Region 64..106
FT /label= Linker_region
FT Domain 107..471
FT /label= catalytic_core_domain
FT /note= "catalytic core domain is the preferred
FT domain for use in constructs of the
FT invention"
XX
XX WO9623928-A1.
XX 08-AUG-1996.
XX
XX 29-JAN-1996; 96WO-US00977.
XX 01-FEB-1995; 95US-0382452.
XX (GEMV ) GENENCOR INT INC.
XX Clarkston KA, Collier KD, Fowler T, Larenas E, Ward M;
XX WPI; 1996-371466/37.
XX N-PSDB; AAT32221.
XX Treatment of cellulose-contg. fabrics such as denim, e.g.
XX stone-washing - using truncated cellulase enzyme to increase
XX abrasion and give reduced redeposition of dye
XX

```

PS Disclosure; Fig 2A-2E; 124pp; English.

XX The amino acid sequences for Trichoderma longibrachiatum
CC cellobiohydrolase I (CBHI) (AAW02022), CBHI1 (AAW02025), endoglucanase I
CC (EgI) (AAW02029), EgII (AAW02032) and EgIII (AAW02034) were deduced from
CC the respective genomic DNA sequences (AAT3220-24). The CBHI, CBHI1,
CC EgI and EgII enzymes have catalytic core domains useful for reducing
CC dye redposition (backstaining) on cellulose-cong. fabrics such as
CC denim, whilst maintaining or increasing abrasion during stonewashing.
CC Truncated enzymes comprising these catalytic core domains can be obt.
CC by proteolysis of the complete enzyme or by inserting the appropriate
CC DNA fragment into a vector, using this to transform a Trichoderma
CC sp. host cell, and recovering the recombinant core domain.

SO Sequence 471 AA;

Query Match 23.9%; Score 538; DB 17; Length 471;
Best Local Similarity 33.3%; Pred. No. 2,1e-37;
Matches 144; Conservative 51; Mismatches 155; Indels 82; Gaps 13;

OY 1 ATHVDNPGATFFVNPYMAQEOSEANQTNATLAKMRVSTYSTAVMND----- 52
DB 110 atysgnptfygtwpanayasevslalpsltgamataaavakvpsftwldtkiplm 169
OY 53 --RIAAINGVNGGGLTYYIDAAISQOOGTPEVIEIYIVDLPGRDCAALASNGELPATA 110
DB 170 egtladitlankng--nyag-----qfvvldpdrdcaalaengeyslad 213
OY 111 AGLOTYETQYIDPIASILSNKYSRLRIYIEPDSLPNATVNNMSIQACATAVPYEEOGI 170
DB 214 ggvakyk-nydltirqlv--veyesdirlllylepdslanlytnlgtckcanaagsyleci 270
OY 171 EYALTKLHAIDNVIYMDAHSGLGMPNNAAGYVQVKVLAASIGVNGIDGFTVTNAN 230
DB 271 nyavtqln-lpnyamyldaghaqylgwpaaqdpaaqlfanvyknaaspralrglatuvan 329
OY 231 YTPLEKPPMTQOVGGQPVESANFYQWNPDIENDAVNDLXSRVLAAGFFSSIGMLIDT 290
DB 330 yn-----gwnltspstytqgnavaynekllylhalqpllanhgw--snaeffltdq 375
OY 291 LRNGMGPNPEPTGPSTATDVNTFVNOSKIDLRHRLGMCNONGAGLGOPPOASPTDFENA 350
DB 376 grsg-----kqptg-----qgqgdgkcnvigtgfgttrpaantgd---s 410
OY 351 HLDAYVWIKPGEESDGTSAASDPTTGKSKDPMCDPTTYSYGVLTNALPNSPIAGQWPPA 410
DB 411 lldsfvwwkpggedgtsdssap---rfdshc-----alpdalqpapqgawfqa 457
OY 411 QFDLVANARPA 422
DB 458 yfvqlltnaaps 469

RESULT 11
ID AAW25789
AAW25789 standard; Protein: 457 AA.

XX AAW25789;
AC
XX
XX 27-MAR-1998 (first entry)
DE Acremonium celluloilyticus cellulase.
KW cellulase; Acremonium celluloilyticus; enzyme.
XX
XX Acremonium celluloilyticus.
OS
FH Key Location/Qualifiers
FT Peptide 1..20
FT Protein /note= "signal peptide"
FT 21..457
FT /note= "mature protein"

XX W09733982-A1.
FN 18-SEP-1997.
XX
PD 14-MAR-1997; 97W0-JP00824.
PF 14-MAR-1997; 96JP-0084479.
XX
PR 14-MAR-1996; 96JP-0084479.
XX
XX (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
PA (MEIJ) MEIJ SEIKA KAISHA LTD.
XX
PI Aoyagi K, Hamaya T, Murakami T, Sumida N, Watanabe M;
PI Yamanobe T;
XX
XX WPI: 1997-470865/43.
DR N-PSDB: AAT91640.
XX
PT Cellulase derived from Acremonium celluloilyticus - also expression
PT vectors used for producing the protein, has improved cellulase
PT activity compared to wild type enzyme
XX
XX Claim 1; Pages 20-24; 47pp; Japanese.
CC This protein with cellulase activity is derived from Acremonium
CC celluloilyticus. Expression vectors containing the encoding DNA sequence
CC can be used to transform microorganisms for the production of this
CC protein. The Acremonium celluloilyticus cellulase has improved cellulase
CC activity compared to the wild type enzyme.

SO Sequence 457 AA;

Query Match 23.7%; Score 534; DB 18; Length 457;
Best Local Similarity 33.8%; Pred. No. 4,5e-37;
Matches 142; Conservative 56; Mismatches 158; Indels 64; Gaps 14;

OY 6 NPVYGATFFVNPYMAQEOSEANQTNATLAKMRVSTYSTAVMNDRIAINGVNGPG 65
DB 99 nftsgyqlyanpysevehltalpsltgslaaatkaelpsfvldtapk-----pt 152
OY 66 LTTYL-DAALSQOOGTPEVEIEI-VIYDLPGRDCAALASNGELPATAAGLOTYETQYIDP 123
DB 153 mgtlylanleaankegasplglftvvydlpdrdcaalaasngeyrvanngyanxa-ylds 211
OY 124 IASILSNKYSRLRIYIEPDSLPNATVNNMSIQACATAVPYEEOGLEYALTKLHAIPNV 183
DB 212 lvaqk--aydvntllllepdslanmwtlnstakcaagsaylcecvyalnln-lanv 268
OY 184 YIYMDAHSGLGMPNNAAGYVQVKVLAASIGVNGIDGFTVTNANTYTPLEKPPMTAQ 243
DB 269 amyldaghaqylgwsanlspaagqlfatvyknaaspastrglatuvanyn-----aw 319
OY 244 QVGGQPVESANFYQWNPDIENDAVNDLXSRVLAAGFFSSIGMLIDTLRNGMGPNPEPTG 303
DB 320 sispp-----stsgdsdydeklylnal脾lltsngwpa-hflmdsting-----vqpt- 369
OY 304 PSTATDVNTFVNOSKIDLRHRLGMCNONGAGLGOPPOASPTDFPNALHLDAYVWIKPGE 363
DB 370 -----kqgdgkcnvigtgfgttrpaantgd---pledaflvwwkpgge 409
OY 364 SDGTSASADPTTGKSKDPMCDPTTYSYGVLTNALPNSPIAGQWPPAOFDQLVANARPAV 423
DB 410 sdtgsnss-----atrydfhcg-----ysdalqpapeagtfwgfayfvqlltnaapal 456

RESULT 12
ID AAY01076
AAY01076 standard; Protein: 446 AA.

XX AAY01076;
AC
XX
XX 08-JUN-1999 (first entry)

XX H: Insolens Cel6B fungal cellulase protein sequence.
 DE
 XX
 XX
 KM Cel6B: fungal cellulase; cleaning composition; conditioner;
 KM cellulolytically active protein; endo-1,4-beta-glucanase; enzyme;
 KM fabric softener; Humicola-like cellulase; glycosyl hydrolase family 6;
 KM detergent composition.
 XX
 OS Humicola insolens.
 XX
 PN MO9901544-A1.
 XX
 PD 14-JAN-1999.
 XX
 PF 02-JUL-1998; 98WO-DK00299.
 XX
 PR 04-JUL-1997; 97DK-0000813.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Andersen KV, Damgaard B, Lund H, Nielsen JB, Schulein M;
 XX
 DR WPI: 1999-106046/09.
 DR N-PSDB: AAX27947, AAX27948.
 XX
 XX
 PT Cleaning composition containing Humicola endo-beta-1,4-glucanase
 PT useful as detergent compositions or additives, or as fabric
 PT conditioners
 XX
 PS Claim 2; Page 247-249; 271pp: English.
 XX
 CC This sequence is the Humicola insolens Cel6B fungal cellulase.
 CC The invention relates to a cleaning composition (A) that contains at
 CC least one enzyme (I) with cellulolytic activity and at least 25 wt.% of
 CC total cellulolytically active protein present is Humicola
 CC endo-1,4-beta-glucanase or a Humicola-like cellulase of the glycosyl
 CC hydrolase family 6. (A) are used as detergent compositions or additives,
 CC or as fabric softener or conditioner. (I) provides colour clarification,
 CC and possibly soil removal, without significant weakening of cellulose
 CC materials during pre-soaking or wet storage (contrast use of enzymes from
 CC families other than 6).
 XX
 SQ Sequence 446 AA:

Query Match 20.0%; Score 449.5; DB 20; Length 446;
 Best Local Similarity 28.0%; Pred. No. 7.2e-30;
 Matches 120; Conservative 70; Mismatches 143; Indels 95; Gaps 16;

QY 6 NPVAGATFFVNPVMAOEVOG-----EAAQTATLAKKRVV-STYSTAVWMDRIAIN 58
 DB 30 npsgrtlvnsdysakldtqrqflsrgdqn---aakkyqekvgtfymnsnflr 86
 QY 59 GVGNGGGLTLYDLAALSQOOGTTPVEVIEIYIDLPERDCAALASNGELPATAAGLOTYET 118
 DB 87 dld-----valqraaekargemp-lyglvlylprdcasagessgelsksgnlrnykn 140
 QY 119 QYIDPFIASISNPKYSLRIVTIIPEDSLPAVATNMSIQACATAPVPEEGIEVALTKLH 178
 DB 141 eynvnpfagklk--aaedvqfavllepdaigmvlygts-afcrnarppqgaigysqlq 197
 QY 179 AIPNVYIYMDAAHSGWLPNNASGVYQVQKVLNASIGVNGIDGFEVTNTANTPLKEPF 238
 DB 198 a-shihlyldvavgvlygvdskleptqevatlqkqgnaklrgfsnvsny----npy 252
 QY 239 MTATQVGGPVSANRYQNNPDIIDENDYAVDLXSLVAAFPSSISMLDLTLANGGCP 298
 DB 253 stamp-----ppytsg-----spspdeeryatnlanamrfglpqf----- 289
 QY 299 NEPTGPSTADVMTFVNSKIDL---ROHGLMCNONGAGIGOPQASPTDFPNAHLDAY 355
 DB 290 -----lldqsrvalagarseswqvcnvpagfgqp---ftctntmnpvdal 332

QY 356 VWIKPEESDGTSAASDPTTGKKSDBPCDPTTYTSGVLTNALPNSPDAQWFPAPFDOL 415
 DB 333 vwwkpgsgsd-----gc-----gnggapaagmwftdayaqml 364
 QY 416 VAAARPAV 423
 DB 365 tqnhdei 372

RESULT 13
 AAM56739
 ID AAM56739 standard; Protein; 449 AA.
 AC AAM56739;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 DE Orpinomyces cellulase celC.
 XX
 KM Cellulase; endoglucanase; cellobiohydrolase; celC.
 XX
 OS Orpinomyces sp. strain PC-2.
 XX
 PN WO9814597-A1.
 XX
 PD 09-APR-1998.
 XX
 PF 03-OCT-1997; 97WO-US18008.
 XX
 PR 04-OCT-1996; 96US-0027883.
 XX
 PA (UYGE-) UNIV GEORGIA RES FOUND INC.
 XX
 PI Chen H, Li X, Ljungdahl LG;
 XX
 DR WPI: 1998-240096/21.
 DR N-PSDB: AAV29472.
 XX
 PT New recombinant DNA encoding Orpinomyces cellulase protein - useful
 PT for, e.g. producing recombinant Orpinomyces cellulase in host cell
 XX
 PS Claim 1; Page 32-34; 69pp: English.
 XX
 CC This polypeptide comprises cellulase celC of Orpinomyces sp. strain
 CC OC-2, an anaerobic bovine rumen fungus Orpinomyces sp. PC-2. Its
 CC amino acid sequence was deduced from an isolated cDNA clone (see
 CC AAV29472). CelC has endoglucanase and cellobiohydrolase activity,
 CC with highest activity at pH 5.6-6.2 and 40 degC. CelA (see AAM56738)
 CC and celB (see AAM56742) cellulolytic enzymes of Orpinomyces sp. PC-2
 CC are also provided. Recombinant DNA molecules encoding Orpinomyces
 CC cellulase proteins are claimed, as well as recombinant cells
 CC selected from Saccharomyces cerevisiae, Escherichia coli,
 CC Aspergillus, Trichoderma reesei, Pichia, Penicillium, Streptomyces
 CC or Bacillus, and a method for producing recombinant cellulase by
 CC culturing these host cells.
 XX
 SQ Sequence 449 AA:

Query Match 19.5%; Score 439.5; DB 19; Length 449;
 Best Local Similarity 32.0%; Pred. No. 5.2e-29;
 Matches 136; Conservative 55; Mismatches 131; Indels 103; Gaps 18;

QY 2 THVDNPVAGATFFVNP-----YMAOEVSAAQNNATLAKKRVVSTYSTAVWMDRIAI 57
 DB 121 tcvnsipstsdnffenealnysnykkfqqevdgsiqrgslqekakvkvypcaavlawsgat 160
 QY 58 NGVNGGGLTLYDLAALSQOOGTTPVEVIEIYIDLPERDCAALASNGELPATAAGLOTYE 117
 DB 181 nev-----arlynags-----kcvfvlylmpirdcnagsgng-----gdndlstyg 223
 QY 118 TQYIDPFIASISNPKYSLRIVTIIPEDSLPAVATNMSIQACATAPVPEEGIEVALTKL 177

Db	224	-gynvnslyntlin--qypnsrlymlipeotlignlvtaann- crnvbmhbxqalsyalskt	279
Oy	178	HAIPNYYIIMDAASHGWLGMPPNNAAGVYQEVOKVLNMSIGVNGIDGFTVTANTYPLKEP	237
Db	280	gtqkuvrvyidaahgwvl---nssadrtaevlaellnagngkrlgystlnvny-----	330
Oy	238	FMNATQGVGGQPRFESANFYQWMPDIDEDYAVNDLXSRLYVAGFFSSJGMLIDTLRNGWG	297
Db	331	-----qpyase--yqyhqnlra-----lstryv-----rgmkflivdsrng---	365
Oy	298	PNEPTGPTATDVTNPTFVNQSKIDLQHRGLMCNONGACIGOPPOASPTDPPNAH-IDAVY	356
Db	366	-----rnpssat-----wcnlkqaglgstrpand--pmpplldayv	400
Oy	357	WIKPGEHSDGTSASDPPTGKSKDPCDPYTTSTYGVLTNALPNSPIAGQWFFAQFDOLV	416
Db	401		442
Oy	417	ANARP 421	
Db	443	enamp 447	
RESULT 14			
AA18414			
ID	AA18414	standard; Protein; 432 AA.	
AC	AA18414;		
DT	15-JAN-2001	(first entry)	
DE	Amino acid sequence of a Celf cellulase protein.		
KW	Celf; cellulase; endoglucanase; cellobiose; glucose; cellobiose;		
KW	cellulosic material; textile; paper; ethanol; fuel;		
OS	Orpinomyces sp.		
XX	Key	Location/Qualifiers	
XX	Peptide	1..21	
FT		/note="signal peptide"	
FT	Protein	22..432	
FT		/note="mature protein"	
FT	Domain	22..57	
FT		/note="fungal cellulose binding domain"	
FT	Peptide	67..105	
FT		/note="Asn-rich linker sequence"	
FT	Domain	106..432	
FT		/note="catalytic domain"	
XX	US6114158-A.		
XX	05-SEP-2000.		
XX	17-JUL-1998;	98US-0118319.	
XX	17-JUL-1998;	98US-0118319.	
XX	(UYGE-) UNIV GEORGIA RES FOUND INC.		
XX	Li X, Ljungdahl LG, Chen H;		
XX	WPI; 2000-593548/56.		
XX	N-PSDB; AAA75403.		
XX	Recombinant DNA molecule encoding a mature Celf cellulase, useful for		
XX	degrading cellulosic material in textile and paper industries, and for		
XX	ethanol production		
XX	Disclosure; Column 13-16; 26pp; English.		
XX	The present sequence represents a Celf protein of Orpinomyces sp. PC-2		

	Cc	CeLF is a cellulase, which has endoglucanase activity and produces
	Cc	cellobiose from cellobiotriose or cellobiotetraose, and glucose and
	Cc	cellobiose from cellobiotriose. The CeLf polynucleotide is useful for
	Cc	recombinantly expressing CeLf mature protein in Escherichia coli or
	Cc	other host cells. The cellulase protein is useful for degrading
	Cc	cellulosic material in textile and paper industries, and also for
	Cc	producing ethanol. Cellulase is also useful for producing fuels and
	Cc	chemicals from cellulosic feed stocks.
SQ		Sequence 432 AA;
XX		
Query Match	19.2%	Score 431.5; DB 21; Length 432;
Best Local Similarity	30.9%;	Pred. No. 2.4e-28;
Matches 132;	Conservativity 66;	Mismatches 126; Indels 103; Gaps 21;
OY	DNPVYGAT-----EFVNPYMAOGVESEAAHQTNATLAAMRVRYSYSTVMNDRIAI	57
	: :: :	:
Db	97 mnsgrsgqnfittnglypnkfleevns-slpklswdlqqkaqvkdvpavwL-----	149
OY	58 NGVNGPG-LTTYLDALSOOQGTTPEVIEIIVIDLPGRCAALASNGELPATAGLOTY	116
	::: :: :	:
Db	150 -awegapgeveqhikaags-----kvvfilymiprdcnsmasag---gsgstinty	197
OY	117 ETOTYDTPIASILSNPKYSLSRLVIITLEPDSLPMNAVTKNSIOACATAVPYEQIGIALFK	176
	: :: :	:
Db	198 k-gyvdnistrlrs--ypnksrvvmlepdltignlvitgsnan-cqnvrqlhknaalsaynv	253
OY	177 LHAIGNVITYMDAHSGLGWPNNNASCYGVGVOKVTLNASICVNG-IDGFVNTANTYPPLK	235
	:: :: :: :: :: :: :: :: :: :: :: :: :: :: ::	:
Db	254 ygaamnnvvvyldaaigkklgyvtclvaav-vkelima--pnkllrglstenvany---	305
OY	236 EPFMTATOOVGQPYESANFYOMWPDIDEADYAVDLVSRLVAAGFSSIGMLIPTLRNGW	295
	::: :: :	:
Db	306 -----qpilas-----eyyhqklasslasavgilp-nmhftivdtgrng-	340
OY	236 GGPNPTGPSTATTDVNTFTVNVNSKIDLROHKGIMCQNAGLGQQPPQASP-TDFPNAHLDA	354
	::: :: :	:
Db	341 -----vdysaatntsc-----twcnfygtfgferprgnpsgmP--llda	378
OY	355 YWMKPGEODGTSAASDPRTGGKSDDPCDPPTYTTSYLVLNALPNSPIAGOWFPAPQEDQ	414
	:: :: :: :: :: :: :: :: :: :: :: :: :: :: ::	:
Db	379 ymwltcpresdgsr-----sgstradpvcs-----rdsllrgdpdqgwThdyfvq	423
OY	415 LVANARP 421	
	::: :: :	:
Db	424 llrnarp 430	
RESULT 15		
AAM56738		
ID AAM56738 standard; Protein; 459 AA.		
NC AAM56738;		
DT 14-SEP-1998 (first entry)		
DE Orpinomyces cellulase cea.		
KX Cellulase; endoglucanase; cellobiohydrolase; cea.		
OS Orpinomyces sp. strain PC-2.		
PN W09814597-A1.		
PD 09-APR-1998.		
PF 03-OCT-1997; 97WO-US18008.		
PR 04-OCT-1996; 96US-0027883.		
RA (UYGE-) UNIV GEORGIA RES FOUND INC.		
XX		

PI Chen H, Li X, Ljungdahl LG:
 XX MPI: 1998-240096/21.
 DR N-PSDB: AAV29471.
 XX
 PT New recombinant DNA encoding Orpinomyces cellulase protein - useful
 for, e.g. producing recombinant Orpinomyces cellulase in host cell
 XX
 PS Claim 1: Page 28-30; 69pp; English.
 XX
 CC This polypeptide comprises cellulase cels of Orpinomyces sp. strain
 CC OC-2, an anaerobic bovine rumen fungus Orpinomyces sp. PC-2. Its
 CC amino acid sequence was deduced from an isolated cDNA clone (see
 CC AAV29471). Cella has endoglucanase and cellobiohydrolase activity,
 CC with highest activity at pH 4.8 and 50 degC. CelB (see AAV56742)
 CC and celC (see AAV56739) cellulolytic enzymes of Orpinomyces sp. PC-2
 CC are also provided. Recombinant DNA molecules encoding Orpinomyces
 CC cellulase proteins are claimed, as well as recombinant cells
 CC selected from Saccharomyces cerevisiae, Escherichia coli,
 CC Aspergillus, Trichoderma reesei, Pichia, penicillium, Streptomyces
 CC or Bacillus, and a method for producing recombinant cellulase by
 CC culturing these host cells.
 XX

Sequence 459 AA:

Query Match 18.3%; Score 411; DB 19; Length 459;
 Best local Similarity 30.0%; Pred. No. 1.5e-26;
 Matches 124; Conservative 66; Mismatches 130; Indels 94; Gaps 20;

OY 13 FEVNPYMAQ-EVQSE---AANQTNATLAAMKRVSTYTAVMMDRIAINGNGP-GLT 67
 DB 133 ffeneIysnykffgveIdIsIkIngdlkakekvkyptavwl-----awdgapevp 185
 OY 68 TYDPAALSQOGGTPPEYIEIVYDLPGRDCAALASNGELPATAAGLOTYETQYIDPIASI 127
 DB 186 ryl-----geagn--ktvfvlymIprcdganasag---gsatldkyk-gylnnlyn- 232
 OY 128 LSNPKYSLRIYVITIEPDSLIPNAVTNMSIQACATAVPEYEGIEYALTKLHAIPNYIYM 187
 DB 233 -tenqykneklvmllepdctlglnvtinn-dncnvrnmhkgalsyslsgfsgshvkvyl 290
 OY 188 DAAHSGMLGMPNNAAGVQEVQKVLNAGISGVNGIDGFTVNTFANTYPLKEPPMTATQOVGG 247
 DB 291 daahgawln--gyadqtaavikeln-naagsgklrglstvnsy----- 331
 OY 248 QPVESANFYQMNPDIIDEADYAVDLYSLVAAGFPSSIGMLIDILRNMGMPNEPTGPRSTA 307
 DB 332 qslse--ykynhqlnra-----leskyv-----rglkflvdtsrng----- 366
 OY 308 TDVNTFEVNOGKIDLRQHGLMCNONGAGLGQPOASPTDPPNAHLDAYVWIKPPGESDGT 367
 DB 367 anvegafnas-----gtwcnfksgaglgqrpkgnpnpsmplldaymwikcpgeadgs 418
 OY 368 SAASDPTGKKSDDPCDPTTYSYGVLTNALPNSPIAGQNFPAQFDQLVANARP 421
 DB 419 s-----qgaradpvc-----argdsIqgapadagswfheyftmlnlganp 457

Search completed: August 29, 2002, 16:21:42
 Job time: 317 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:24:01 ; Search time 72.78 Seconds
(without alignments)
558,475 Million cell updates/sec

Title: US-09-917-384-4

Perfect score: 2249
Sequence: 1 ATHVDNRYAGATFEVNPYWA.....AGQWFPAPQFDLVANARPAY 423

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1244	55.3	872	2	S49541
2	1171	52.1	683	2	A82704
3	1050	46.7	579	2	T35240
4	999	44.4	596	2	A55976
5	639.5	28.4	438	1	S70602
6	546	24.3	471	1	A26160
7	546	24.3	471	1	A38979
8	330.5	14.7	380	2	G70847
9	261.5	11.6	441	2	T12011
10	244.5	10.9	449	2	A24993
11	242.5	10.8	426	2	A42360
12	193.5	8.6	359	2	JN0544
13	143	6.4	388	2	J70308
14	120	5.3	1180	2	E86719
15	113.5	5.0	693	2	D90441
16	113	5.0	638	2	G97171
17	112	5.0	329	2	G82900
18	111.5	5.0	1131	2	T41144
19	110	4.9	2348	2	AD1841
20	108.5	4.8	745	2	G72453
21	107	4.8	657	1	A64079
22	107	4.8	728	1	A38084
23	105.5	4.7	1217	1	G1BPT4
24	103	4.6	583	2	G90327
25	103	4.6	1268	2	B99789
26	103	4.6	1270	2	E85649
27	103	4.6	1270	2	E85649
28	103	4.6	3029	2	S76109
29	103	4.6	3972	2	S75251

30	102.5	4.6	691	2	T44543	probable bacteriophage
31	102.5	4.6	1441	2	B86807	hypothetical prote
32	101.5	4.5	3191	2	T22945	hypothetical prote
33	100	4.4	475	2	AD3287	zinc metalloprote
34	99.5	4.4	591	2	S49829	F420-nonreducting h
35	99.5	4.4	878	2	A96969	cellulase CelE ort
36	99	4.4	385	2	H70503	probable ppe prote
37	99	4.4	588	2	A84108	N-acetylmuramoyl-L
38	99	4.4	739	2	S15727	cellulase (EC 3.2.
39	99	4.4	2124	2	A28452	proteoglycan core
40	98.5	4.4	512	2	F85342	beta-1, 3-glucanas
41	98.5	4.4	512	2	S50599	SHC1 protein - yea
42	98.5	4.4	1136	1	USB581	parasporeal crystal
43	98.5	4.4	1474	2	F69009	probable membrane
44	98	4.4	502	2	A48679	differentiated ker
45	98	4.4	657	2	AE0431	probable exported

ALIGNMENTS

RESULT 1
S49541
cellulase - Cellulomonas flm1
C:Species: Cellulomonas flm1
C:Date: 29-Nov-1995 #sequence_revision 01-Mar-1996 #text_change 22-Oct-1999
C:Accession: S49541; A47093
R:Meinke, A.; Gilkes, N.R.; Kibbun, D.G.; Warren, R.A.J.; Miller Jr., R.C.
MOL. Microbiol. 12, 413-422, 1994
A:Title: Cellulobiohydrolase A (CbhA) from the cellulolytic bacterium Cellulomonas flm1
A:Reference number: S49541; M01D:94344030
A:Accession: S49541
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-872 <MB1>
A:Cross-references: EMBL:125809; NID:9456028; PIDN:AC36898.1; PID:9456029
R:Meinke, A.; Gilkes, N.R.; Kibbun, D.G.; Miller Jr., R.C.; Warren, R.A.
J. Bacteriol. 175, 1910-1918, 1993
A:Title: Cellulose-binding polypeptides from Cellulomonas flm1: endoglucanase D (Cend
A:Reference number: A47093; M01D:93209933
A:Accession: A47093
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 812-872 <MB2>
A:Experimental source: ATCC 484
A>Note: sequence extracted from NCBI backbone (NCBIN:128120, NCBIR:128121)
C:Superfamily: Bacterial cellulose-binding domain homology
F:769-870/Domain: bacterial cellulose-binding domain homology <BCB>
F:770-869/Disulfide bonds: #status predicted

Query Match	55.3%	Score 1244	DB 2	Length 872
Best Local Similarity	55.7%	Pred. No. 2.1e-63		
Matches 239	Conservative 60	Mismatches 120	Indels 10	Gaps 6
QY	3	HYNDPAGATFEVNPYAOESEAQNT-NATLAARVSTVAVMMDRIAINGVN 61		
DB	44	HYNDPAGATFEVNPYAOESEAQNT-NATLAARVSTVAVMMDRIAINGVN 103		
QY	62	GGPGLTYYDAALSOOGT-TPEVIEIYVDLPGRCDAALASNGELPATAAGLOTYETQY 120		
DB	104	DGNGLRKHLDNAVAQKACAGVPLVFNLYVDLPGRCPALASNGELPATAAGLOTYETQY 163		
QY	121	IDPFIASLSPKYSLSLIVITIEDSPNNAVTKNSIOACATVYVYOGIEYATLTKLHAI 180		
DB	164	IDPFIADLDDPEYESSIRIAATTEPDSLPLNTTINISEPACCOAAFYVQGVKVALDKLHAI 223		
QY	181	PNYIYVDAASGWLGPNNNA---SGYVOEYOKVLNASIGVNGIDGFTVTANTPTLKEP 237		
DB	224	PNYIYVDAASGWLGPNNNA---SGYVOEYOKVLNASIGVNGIDGFTVTANTPTLKEP 281		
QY	238	FMT-ATQOVGQPVESANFYQWMPDIDEDYAVDLYSLVAAGFPSSIGMLIDTLRNGWG 296		

Db 282 ILSDSLINNTPIRSKAFYEMNFDDEIDYTAHMHRLVLAAGFPSSIGMLVDSRNGW 341

QY 297 GPNBETGSTATDVNTFYNQSKIDLRQRHGLMCNONGAGLQPPQASPTDFPNAHLDAY 356

Db 342 GPNRPTSTTASTDVAAYADNARVDRHRCAMCDLGAIGIRPEDEAFSPGSAASHLDAFV 401

QY 357 WIKPGEEDGSAASADPTTGKSDMCDPTYTTS--YVULNALPNNSPIAOWPFAODQ 414

Db 402 WIKPGEEDGSAADPTIPNDQGRFDRMCDPTFVSPKLNOLTGATPNAFLAQWFEEDPVT 461

QY 415 LVANARPAV 423

Db 462 LVKANAPVT 470

RESULT 2

A82704

1.4-beta-cellobiosidase XF1267 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: A82704

R:anonymous, 'The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequence

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: A82704

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-683 <SIM>

A:Cross-references: GB:AE003960; GB:AE003849; NID:9106242; PIDN:MAF84076.1; GSPDB:GN001

A:Experimental source: Strain 9a5c

R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

R:Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre

as-Neto, E.; Docena, C.; El-Dorri, H.; Facinani, A.P.; Ferreira, A.J.S.

submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm

J.D.; Jungueira, M.L.; Kemper, E.L.; Kitzima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, B

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava

M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; V

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF1267

Query Match 52.1%; Score 1171; DB 2; Length 683;

Best Local Similarity 53.0%; Pred. No. 3,46-78;

Matches 227; Conservative 59; Mismatches 120; Indels 12; Gaps 6;

QY 3 HYDNFYAGTFFVNPVMAQVOSEAPANOTNATLAAKRFVSTYTAVMMDRIAALINGVNG 62

Db 21 HYDNFVATATSLANDYKSDIDTSLAKVNDVSLKMKMOTIKNYPGWMLDRIIDAKGSGPK 80

QY 63 GGC---LTTTVDAAALSOOGGTPPEVETLEIYIDLPDRCDAALASNGELPATAAGLOTYETQ 119

Db 81 NFGRLNTLGHDLAALQKKNTPTATVYIYDIPERDCHALASNGELPTEBGLDQRYRRE 140

QY 120 YIDPASTILSNPKVSSLSLIVTILPEPDSLPNAVNTNNSIOACATA--VPYYEODIGEVALTKL 177

Db 141 YIDTILASIFSNSKRYKIDIVNIIEPDSLPNIITNSTPSCDAKMTGIEGCIKALKKL 200

QY 178 HAIPNVYIYMDAASHGMLGWPNNASGVYQVOKVL--NASIGVNGIDGFVTNTANTPLKE 236

Db 201 SEIPNVYNYMDIGSHGMLGMDTNRIPAVSLTYKVIOSITAGPASYNGCATVNTANTPLIE 260

QY 237 PPM-TATQOVGGQPVESANFYQWMDIDEDADYADVLSRLVAAGFPSSIGMLIDPLRWGW 295

Db 261 PULPMDINIGGQPIRSSKRYEWMNRYFDEMDYSETLYNDYFAAAGFPSSIGFIDGRNGW 320

QY 296 GGPNEPTSPATVATVNFVNSOKIDLRHNRGLMCNQNAGSGAPQOASPTDFEPNANHAY 355

Db 321 GGPERFTS-AGCNDVNSIVNSGRIDRRNRHRCNMCQKQAGIGLPIAT- ---PGSHDAF 375

QY 356 WIKRPGESDGTSAASDPTTGKSKSDPMCDPTTYSYGLVTNALPNSPIAGOWFPAQFDOL 415

Db 376 OMKRPGVSDGSSSLIPNDGCKGFRDYCDPFTTTPDVLGALPDAPLISGDMFNAQFKL 435

QY 416 VAMARPAV 423

Db 436 INNAYPD1 443

RESULT 3

T35240

probable secreted cellulase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

C:Accession: T35240

R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, September 1998

A:Reference number: Z21572

A:Accession: T35240

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-579 <SEP>

A:Cross-references: EMBL:AL031515; PIRN:CAA20645.1; GSPDB:GN00070; SCOEDB:SC5C7.33

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC5C7.33

Query Match 46.7%; Score 1050; DB 2; Length 579;

Best Local Similarity 49.8%; Pred. No. 2,1e-69;

Matches 214; Conservative 61; Mismatches 125; Indels 30; Gaps 8;

QY 4 VDNPPAGATPFNPPYMAQEOVSEANQTNATLAAMRVSTYSTAVMMDRTAINGVNG 63

Db 165 VDNPPDGAQVYVNPBMSANAAAEFGDR-----IADEPTGVWLDRTAALIGANGG 214

QY 64 PGLTYYDADALSQQGGTTPEVIEIYIDLPGRDCALASNGELPTAAGLQTYETQYIDP 123

Db 215 MGLRHDLDAL- EKGSGEMVQVLYIYMLPGRDCSALASNGELGPT- -IDRYKTEYIDP 271

QY 124 IASTLSNKKYSLSKRVTLIEPDSLPNATYNNISQACAF-----AVYIEDEGIEYALTK 176

Db 272 IAEILSDSKYADLRIVTVEIDSLPMLTYNVSGRPTAENCDDVMKANQNYQKGVGYALNR 331

QY 177 LHAIPNVYIYMDAAHSGMLGMPNNASGVYOEQVKLNASIG- -VNGIDGFYNTANTYPL 234

Db 332 LGAVGNVYNYVDAGHGMLGWDN- FGSALIEFKTAATTEGATLDDVGFIVNTANYAL 390

QY 235 KEPEMTATQVGGQGVESANFYQWNPDIIDEADYADVLSRLVAAGFPSSIGMLIDTLRNG 294

Db 391 KEENFKIDTSVNGTSVRSDWVWQNYDELSTYAOAMRDKLISLGFQDNLGLIDTSRNG 450

QY 295 WGGPRHEPTGPSTATVNFVNSKIDLRQHNGLMCNQAGACIGQRPQASPHDFPAHADA 354

Db 451 WGGADRPYGPATIDVNTYVNGGRRDRIRILHGMNCSGAGIGERPOASPA---AGIDA 506

QY 355 YWIKRPGESDGTSAASDPTTGKSKSDPMCDPTT- ---SYGLVTNALPNSPIAGOWFPAQ 411

Db 507 YWMMKRPESDSSSKLIDNPDKGFRDRCDDPTTYTENRGNMSGALLDAPLISGAMFSAQ 566

QY 412 FDOULVANNAP 421

Db 567 FOELMKRNAYP 576

RESULT 4

A55976

cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) - Thermomonospora fusca

C:Species: Thermomonospora fusca
 C:date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 22-Oct-1999
 C:Accession: A55976
 R:Zhang, S.; Lao, G.; Wilson, D.B.
 Biochemistry 34, 3386-3395, 1995
 A:title: Characterization of a Thermomonospora fusca exocellulase.
 A:Reference number: A55976; MUID:95186496
 A:Accession: A55976
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-596 <SHA>
 A:Cross-references: GB:U18978; NID:9664822; PIDN:AAA6221.1; PID:9664823
 C:Superfamily: bacterial cellulose-binding domain homology
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:40-140/Domain: bacterial cellulose-binding domain homology <CB>
 F:41-139/Disulfide bonds: #status predicted

Query Match 44.4%; Score 999; DB 2; Length 596;
 Best Local Similarity 48.3%; Pred. No. 1.2e-65;
 Matches 209; Conservative 55; Mismatches 137; Indels 32; Gaps 9;

QY 4 VDNVYAGATFEVNPYVAQVQSEANQNTNATLAKMRVYSTSTAVMMRIALING---- 59
 DB 179 VDNVPGAKLYNPVWSAKAAEPGGS-----AVANESTAVWLDRIIGALIEGNDSP 228
 QY 60 VNGGPELTYLDAALSQOQGTPEVIEIYIDLPGRDCALASNGELPATAGIQETQ 119
 DB 229 TTGSMKLRHLEAV-RQSGDPLTIQVYINLPGRDCALASNGELGPE--LDRKSE 265
 QY 120 YIDPIASILSN-KRYSRLVITLIEPDSLPAVNTNMSIQACATVPY-----YEOGIEY 172
 DB 286 YIDPIADIMDFADYEELRIVALIIEIDSLPNTLVNNGNGCTELCAVMKONGVYNGV 345
 QY 173 ALTKLAIPVNYLYMAAHSGMLGMPNNSAGYQVQKVLNAS-IGVNGIDEGVTRTANY 231
 DB 346 ALTKLEIPVNYIDAHHGMLGMDSNFEGPSVDFEAMNAGSTVDYVHGISNANY 405
 QY 232 TPLKEPMTATQVGVGPVSEANFQWNPDIADYADVADLYSRVAAFPSSIGMLIDL 291
 DB 406 SATVEEYLDVNGVNTQNLIRQSKWVDMNQYVDELSPQDLRQALIAKGFSDIGMLIDTS 465
 QY 292 RNCGSPNEPTGPSTATDVNTFYVNSKIDLRQHRGLMCNONGAGLQPPQASPTDFPNH 351
 DB 466 RNCMGSPNPTGPSSSTDLTYVDESRIIDRRIRPGMNCNAGAGLGERPTVNSA---PG 521
 QY 352 LDYVYAIKPGESDGTSAASDPTTKKSDPMCDPTTTSYGVLTN---ALPNPIAGQWF 408
 DB 522 VDYVYVWKPGESDGASEEIPNDEGKGFDRMCDPTTQGNARNGNPSGALPNAPISGHWF 581
 QY 409 PAQFDQLVANARP 421
 DB 582 SAQFRELLANARP 594

RESULT 5
 S70602
 cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) II precursor - cultivated mushroom
 N:Alternate names: cellulase
 C:Species: Agaricus bisporus (cultivated mushroom)
 C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C:Accession: S70602
 R:Jaeger, E.; Chow, C.M.; Challen, M.P.; Thurston, C.F.
 Curr. Genet. 30, 56-61, 1996
 A:title: Correlation of exons with functional domains and folding regions in a cellulase
 A:Reference number: S70602; MUID:96269930
 A:Accession: S70602
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-438 <YAG>
 A:Cross-references: EMBL:Z34007; NID:g1494969; PIDN:CAAB3971.1; PID:g1494970
 C:Genetics: 34/1; 56/2; 96/3; 202/1; 325/3; 410/2

C:Superfamily: cellulose 1,4-beta-cellulobiosidase II; fungal cellulose-binding domain
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:25-56/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 28.4%; Score 639.5; DB 1; Length 438;
 Best Local Similarity 36.6%; Pred. No. 2.2e-39;
 Matches 153; Conservative 62; Mismatches 134; Indels 69; Gaps 12;

QY 6 NPVYAGATFEVNPYVAQVQSEANQNTNATLAKMRVYSTSTAVMMRIALINGVNGPG 65
 DB 90 NPYTGKTWLSPEYADVEAQAADISNPSTAKASVAKIPTFEWFTVAV-----PD 143
 QY 66 LTYLDAALSQOQGTPEVIEIYIDLPGRDCALASNGELPATAGIQETQYIDPIA 125
 DB 144 LGGYLAARSKNQ-----LVQIVYDLPDQDCALASNGEESLANDLNTK-NYVQIA 197
 QY 126 SILSNPKYSSLRVITLIEPDSLPAVNTNMSIQACATVPYEOGIEYALTKLAIPVNY 185
 DB 198 AOIK--QPPVSYVAVIEPDSLPAVNTLVNVOKCANASAKKEGVYAVOKLNAV-GVTM 254
 QY 186 YMAAHSGMLGMPNNSAGYQVQKVLNASIGVNGIDGFTVNTFNTPLKEPMTATQV 245
 DB 255 YIDGHAQGLGMPANLSPAAQLERQIYRDAGSPNRLGIALTNVANFNALRAS----- 306
 QY 246 GGPVSEANFQWNPDIADYADVADLYSRVAAFPSSIGMLIDLNCGMPNEPTGPS 305
 DB 307 SPDPIT-----QGSNNDLHITIALAPMLNSMGFPNH--PIYQGSNG----- 348
 QY 306 TATDVNTFYVNSKIDLRQHRGLMCNONGAGLQPPQASPTDFPNHLDYVYAIKPGESD 365
 DB 349 -----VONIRDMQDMCNVAGCGQR--TTWTGSSLIDAIYVWVPGGEC 393
 QY 366 GTSAASDPTTKKSDPMCDPTTTSYGVLTNALPNSEIAGWFPQADQLVANARPV 423
 DB 394 GTSNNSP-----RDSHCS-----LSDAQPAEAGTWQVETTVANANPAL 438

RESULT 6
 A26160
 cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) II precursor - fungus (Trichoderma reesei)
 N:Alternate names: exo-cellulobiohydrolase II
 C:Species: Trichoderma reesei
 C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A26160
 R:Chen, C.M.; Gritzali, M.; Stafford, D.W.
 Bio/Technology 5, 274-278, 1987
 A:title: Nucleotide sequence and deduced primary structure of cellobiohydrolase II fr
 A:Reference number: A26160
 A:Accession: A26160
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-471 <CHE>
 R:Teeri, T.T.; Lehtovaara, P.; Kauppinen, S.; Salovuori, I.; Knowles, J.
 Gene 51, 43-52, 1987
 A:title: Homologous domains in Trichoderma reesei cellulolytic enzymes: gene sequence
 A:Reference number: A26472; MUID:87248061
 A:Accession: A26472
 A:Molecule type: DNA
 A:Residues: 1-471 <TEE>
 A:Cross-references: GB:M16190; NID:g170540; PIDN:AAA34210.1; PID:g170541
 C:Genetics: 31/2; 160/2; 243/1
 C:Superfamily: cellulose 1,4-beta-cellulobiosidase II; fungal cellulose-binding domain
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:31-62/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 24.3%; Score 546; DB 1; Length 471;
 Best Local Similarity 33.6%; Pred. No. 1.9e-32;
 Matches 145; Conservative 51; Mismatches 154; Indels 82; Gaps 13;
 QY 1 ATHVDNYPYAGATFEVNPYVAQVQSEANQNTNATLAKMRVYSTSTAVMM----- 52

```
Db 110 ATTSNPFVGTTPMANAYTASVSSALPISLGAAMATAAAAKVPSFWMULTDTPM 169
QY 53 --RIAINGVNGGPGITTYLDAALSQOOGTPEVIEIYIDLPGRDCAALASNGELPATA 110
Db 170 EOTLADIRFANKNG--NYAG-----QFVYDLPDRDCAALASNGEYSTAD 213
QY 111 AGLOTIEQYIDPIASISNPKYSRLVTIIEPDSLPAVATNMSIQACATVPEYEGCI 170
Db 214 GGVAYKK-NYIDTIRQIY--VEYSDIRLLVIEPDSLANTLTNLTGTPCANQASVYLCFI 270
QY 171 EYALTKLAIPNVYTYMDAAHSGMLGMPNNAAGYQVEQVKYLNASIGVNGIDGFYNTAN 230
Db 271 NYAVYQLN-LRVVAMYTLDAAGHAGMLGMPANODPAQOLFANYKNAASSPRLALGLATNVAN 329
QY 231 YTPLEKPEMTATQOYGQPVESANFYQWNPDIIDEADYAVDLYSRVLAAGFPSSIGMLIDT 290
Db 330 YN-----GWNITSPSYTQGNNAVYNEKLYTHAIGPLLANNGW-SNAFFITDQ 375
QY 291 LRNGGNGNEPTGPSTADVNTFVNGSKIDLRQHRGLMCNONGAGLGPPOASPTDFPNA 350
Db 376 GRSG-----KQPTG-----QOQWGMCMNVGTGFGIRPSANTGD---S 410
QY 351 HUDAYVWIKPGEESDGTSAADPTTGKSDPCDPTTYSYGLTNALPNSPIAGQWPPA 410
Db 411 LLDSEVWVKPGECDGTSSSAP-----RPSHC-----ALPDALQAPQAQAMFOA 457
QY 411 QPDQVANARPA 422
Db 458 YFVOLLTNANPS 469
```

RESULT 7

cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) II - fungus (*Trichoderma viride*)
C:Species: *Trichoderma viride*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A38979
R:Wang, J.R.; Zhang, M.F.; Huang, T.
Acta Genet. Sin. 22, 74-80, 1995
A:Title: The primary structure of cellulohydrazase gene (CBH II) from *Trichoderma viride*.
A:Reference number: A38979
A:Accession: A38979
A:Molecule type: DNA
A:Residues: 1-471 <MAN>
C:Genetics:
A:Gene: cbhII
A:Introns: 31/2; 160/2; 243/1
C:Superfamily: cellulose 1,4-beta-cellulobiosidase II; fungal cellulose-binding domain hom
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:31-62/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 24.3%; Score 546; DB 1; Length 471;
Best Local Similarity 33.6%; Pred. No. 1.9e-32;
Matches 145; Conservative 51; Mismatches 154; Indels 82; Gaps 13;

```
QY 1 ATHDNPYAGATFEVNPYMAQEOSEANQTNATLAARKRVSTYSTAVMD----- 52
Db 110 ATYSGNPFVGTTPMANAYTASVSSALPISLGAAMATAAAAKVPSFWMULTDTPM 169
QY 53 --RIAINGVNGGPGITTYLDAALSQOOGTPEVIEIYIDLPGRDCAALASNGELPATA 110
Db 170 EOTLADIRFANKNG--NYAG-----QFVYDLPDRDCAALASNGEYSTAD 213
QY 111 AGLOTIEQYIDPIASISNPKYSRLVTIIEPDSLPAVATNMSIQACATVPEYEGCI 170
Db 214 GGVAYKK-NYIDTIRQIY--VEYSDIRLLVIEPDSLANTLTNLTGTPCANQASVYLCFI 270
QY 171 EYALTKLAIPNVYTYMDAAHSGMLGMPNNAAGYQVEQVKYLNASIGVNGIDGFYNTAN 230
Db 271 NYAVYQLN-LRVVAMYTLDAAGHAGMLGMPANODPAQOLFANYKNAASSPRLALGLATNVAN 329
```

```
QY 231 YTPLEKPEMTATQOYGQPVESANFYQWNPDIIDEADYAVDLYSRVLAAGFPSSIGMLIDT 290
Db 330 YN-----GWNITSPSYTQGNNAVYNEKLYTHAIGPLLANNGW-SNAFFITDQ 375
QY 291 LRNGGNGNEPTGPSTADVNTFVNGSKIDLRQHRGLMCNONGAGLGPPOASPTDFPNA 350
Db 376 GRSG-----KQPTG-----QOQWGMCMNVGTGFGIRPSANTGD---S 410
QY 351 HUDAYVWIKPGEESDGTSAADPTTGKSDPCDPTTYSYGLTNALPNSPIAGQWPPA 410
Db 411 LLDSEVWVKPGECDGTSSSAP-----RPSHC-----ALPDALQAPQAQAMFOA 457
QY 411 QPDQVANARPA 422
Db 458 YFVOLLTNANPS 469
```

RESULT 8

probable cellulase - *Mycobacterium tuberculosis* (strain H37RV)
C:Species: *Mycobacterium tuberculosis*
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70847
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: G70847
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Residues: 1-380 <COL>
A:Molecule type: DNA
A:Cross-references: GB:A021428; GB:AL123456; NID:g3261514; PIDN:CAA16243.1; PID:e124
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: celA

Query Match 14.7%; Score 330.5; DB 2; Length 380;
Best Local Similarity 29.1%; Pred. No. 9.7e-17;
Matches 109; Conservative 42; Mismatches 123; Indels 101; Gaps 18;

```
QY 6 NPYAGATFEVNPYMAQEOSEANQTNATLAARKRVSTYSTAVMDRIAINGVNGPG 65
Db 89 NPLAKPFYVDPAASAMVAAARNANPNMELTGVANTPOSY----WLDQ--APPATVGT 142
QY 66 LTTIYDAALSQOOGTPEVIEIYIDLPGRDCAALASNGELPATAAGLOTIEQYIDPIA 125
Db 143 VARYTGAA--QAAAMP--VLTJGIRPHRDCGSYASG--FATG--TDYRGWIDAVA 191
QY 126 SIL-SNPKYSRLVTIIEPDSLPAVATNMSIQACATVPEYEG--LEYALTKLAIPN 182
Db 192 SGLGSSP-----ATLYEPDL-----AMADLSPOQORERDLYAVDTLTRDPA 238
QY 183 YTYMDAAHSGMLGMPNNAAGYQVEQVKYLNASIGVNGIDGFYNTANTPLKEPKMTAT 242
Db 239 AAVYVDAGHSRLS-----AEAMAAARLN-DVGVGARGRSLVSN-----FYTTD 282
QY 243 QOYGQPVESANFYQWNPDIIDEADYAVDLYSRVLAAGFPSSIGMLIDTLRNGGNGPNEPT 302
Db 283 EETG-----YGEAT-SGLTNGSHYVIDTSRNG-----A 309
QY 303 GPSTADVNTFVNGSKIDLRQHRGLMCNONGAGLGPPOASPTDPNHLADYVWIKPPG 362
Db 310 GPAAPAPLN-----WCNPSGRALGAPP--TTATAGAHADATYVWIKRPG 350
QY 363 ESDGTSAADPTTGK 377
Db 351 ESDGTGGRGEPQAGR 365
```


Db 310 -----RTVTRCPG--VDAPLMT-----TCPVTDGGDGPVFSPPKL----- 342
 Qy 394 LTNALPNSPIACQWPPAPDOLVANAR 420
 Db 343 ---QLPRKPAAGR---GCRDTIVRSAR 363

RESULT 14

E86719

hypothetical protein yghd [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001

C:Accession: E86719

R:Boletín, A.; Winkler, P.; Manger, S.; Jallón, O.; Malarme, K.; Weissenbach, J.; Ehrlich

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss

A:Reference number: A86625; MUID:21235186; PMID:11337471

A:Accession: E86719

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1180 <STO>

A:Cross-references: GB:AE005176; PID:g12723674; PIDN:AAK04855.1; GSPDB:GN00146

C:Genetics:

A:Gene: yghd

Query Match 5.3%; Score 120; DB 2; Length 1180;
 Best Local Similarity 21.4%; Pred. No. 1.6; Mismatches 155; Indels 178; Gaps 26;
 Matches 105; Conservative 53;

Qy 6 NPYAGATFEVNPYMAOEVOSEANOTNATLAKMRVSTY---STAVMDRIAAI--- 57
 Db 579 NPVYCATSTLNTY---TVTLTAVSQT-----ATYTGVTSTALNTPAIPALIS 624
 Qy 58 NGVNGPGLT-----YLD-----ALSQOQT----- 80
 Db 625 TGTGG-SVTTPTGYTSPQIPSGYISAIYAGTSASGTLISSSTNGTNMAAPSGQVYA 683
 Qy 81 -----TPEVIRIV-----YDLPGRDCALASNGELPATAGL-----QTYEYOYI 121
 Db 684 PTGNQYIOLAPRTVAVTFISIDPNPNDPLNVAOGGTPYVGSILGSGTTNNYOSQYI 743
 Qy 122 DPASILSNPKYSSLRIVTIIIEPDSLPMNAVTNMSIQACATAPVY-----YEOGIEYALT 175
 Db 744 DMNNSLTGN-----TLSTTD-----NVTYDGAATGKIYIVTGYOYOGCTTTTAT 787
 Qy 176 KLHAIVNYIYD--AAHSGML-GMPNNAAGYVQEVQKVLNASIGVNGIDGFTYNTA-NY 231
 Db 788 NSAPITTYKTFSDLAHAYVLAGTSNRTOLVHLMTVDQTTITSNSTSGQHVNPATNY 847
 Qy 232 TPLKEPFMTATQOVGQPVESANFYQWNPDIIDADYAVDLYSRLVAAGPSSIGMLIDL 291
 Db 848 IFSSE--ITSSYDIDG-----ADDELSTAINGNATV----- 878
 Qy 292 RKGWGGPNPEPTGPSTATDVNTFYNSKIDLRQHRGLMCNQNGAGLQPPQASPTDEPNMH 351
 Db 879 -NITGGQTOYWDAT-----NSQLSIGTYTENYVALNTYG----- 913
 Qy 352 LDAYV-WIRPGES-----DGTSAASDPTTGKKSDEPCDPT--YTSYGVLTNA 397
 Db 914 VQAYTWM--PSHSGTQSOFLOTLDATSLANDTVTSQFTLEVDQATATYTNHNIIPRA 971
 Qy 398 L--PNSPIAG 405
 Db 972 ASYNPSSDLVG 982

RESULT 15

D90441

ABC transporter, periplasmic component SSO2669 [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C:Accession: D90441
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder
 aretti, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.
 submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: D90441

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-693 <KUR>

A:Cross-references: GB:AE006641; NID:g13815984; PIDN:AAK42787.1; GSPDB:GN00155

C:Genetics:

A:Gene: SSO2669

Query Match 5.0%; Score 113.5; DB 2; Length 693;
 Best Local Similarity 22.1%; Pred. No. 2.2; Mismatches 190; Indels 107; Gaps 26;
 Matches 102; Conservative 63;

Qy 7 PYAGATFEVNPYMAOEVOSE--AANOTNATLAK-----NRVSTYSTAVMDRIAAING 59
 Db 242 PYSG-----PFLSOLSNHEIYLNKPNRYDAKIYFNOVITTYOGQA---DLAA--A 230
 Qy 60 VNGG-----PGLTLYDLAALSOQO--GTPPEVIEIYDLPGRDCALASNGELPATAG 112
 Db 291 ITGGQVSMDSGLT-----GLSPQLESPLPTLEVI--NLPQFPGMGIAFNLNKPN---W 339
 Qy 113 LQYEVQYIDPISILSNPKYSSLRIVTIIIEPDSLPMNAVTNMSIQACATAPVYEOGIEY 172
 Db 340 LREYQVR--AAIAYILNRATIASVG--GPLTAPVTIPNAPNLISYSPMTSSQYSSLNPPY 396
 Qy 173 --ALTK-LHAIPNVYIYMDAAHSGMLG-----PNNAGYVOEVQKVL 212
 Db 397 NVNLTKAAQLKSVGYQKSG-----VWYTPNCTPPLITIGASPPFOALAMHEVOKEL 451
 Qy 213 NASIGN-----GIDGFTN--TANTPLKEPFMTATQOVGQPV 251
 Db 452 Q-QFGINVOLHIYTVVSOHWQAMONGGYDLMFENNGSSVPGTAPWSLVLSYFGYP-- 508
 Qy 252 SANFYQWNPDIIDAD--YAVDLYSRLVAAGPSSIGMLIDLKNGWGPNE--PTGPSTA 307
 Db 509 -WNVQWNNENTLPLNGTIIIDFHLLEETESPNTQOLIQANDELSYMMOYLLPILPLVE 567
 Qy 308 TDVNTFYNSKIDLRQHRGLMCNQNGA--GLGQPPQASPTDFPNAHLDAYVWIRPGES-- 364
 Db 568 IENVYIVNPSLLTAAPRANSMWEBAQYIG-----GTAMVALID--YWAPLYESTI 619
 Qy 365 --DGTSAASDPTTGKKSDEPCDPTTYTSYGVLTNALPNSPI 403
 Db 620 ITTTTTSISTTITTTTTSATATATSVTTTTSVTTTTSISTTIV 661

Search completed: August 29, 2002, 16:24:03
 Job time: 338 sec

...

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:38:31 ; Search time 38.44 Seconds
(without alignments)
426.076 Million cell updates/sec

Title: US-09-917-384-4

Perfect score: 2249
Sequence: 1 ATHVDNPPYAGATFFVNPYMA.....AGQWFPAGFDQLVANRPAY 423

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1244	55.3	872	1	GUXX_CELFI
2	639.5	28.4	438	1	GUXX_AGABI
3	358	24.8	462	1	GUXX_FUSOX
4	346	24.3	471	1	GUXX_TIRRE
5	261.5	11.6	441	1	GUXX_THERF
6	244.5	10.9	449	1	GUXX_CELFI
7	238.5	10.6	456	1	GUXX_MCHT
8	193.5	8.6	359	1	GUXX_STRSQ
9	181.5	8.1	321	1	GUXX_STRHA
10	112	5.0	790	1	QUXA_XANCT
11	107	4.8	657	1	CNI6_HAEIN
12	107	4.8	657	1	GAOA_DACDE
13	103.5	4.6	527	1	VG12_BPT4
14	100.5	4.5	321	1	EL3F_HORVU
15	99	4.4	739	1	GUXX_CLOTM
16	99	4.4	2124	1	PGCA_RAT
17	98.5	4.4	443	1	HXA3_MOUSE
18	98.5	4.4	512	1	YES6_YEAST
19	98.5	4.4	1136	1	CABA_BACTI
20	98.5	4.4	1567	1	PMN2_MOUSE
21	98	4.4	423	1	PM5_MGCPN
22	97.5	4.3	561	1	LCFA_ECOLI
23	97	4.3	1005	1	Y456_CHLTR
24	96	4.3	3354	1	CADN_HUMAN
25	96	4.3	910	1	DNJM_MYCPN
26	96	4.3	1196	1	AMTB_PAEPO
27	95.5	4.2	474	1	ENCB_HUMAN
28	95	4.2	433	1	ENO_BORBO
29	95	4.2	2132	1	PGCA_MOUSE
30	94.5	4.2	1589	1	PHP_DROME
31	94.5	4.2	1742	1	GUXX_CALSA
32	94	4.2	561	1	HLVB_PROMT
33	94	4.2	625	1	AMTG_NEUCR

34	94	4.2	768	1	AMT1_SACDI	P29160 saccharomyc
35	94	4.2	936	1	CDAS_HUMAN	Q95177 homo sapien
36	93.5	4.2	294	1	MTAI_SOREQ	O13595 soraria eq
37	93.5	4.2	837	1	LDLR_RABIT	P20063 oryctolagus
38	93	4.1	530	1	HLX_HA117	P29143 halophilic
39	93	4.1	809	1	COAT_DSDNV	Q71155 diatraea sa
40	93	4.1	937	1	CDAY_HUMAN	Q9UN72 homo sapien
41	93	4.1	1134	1	YML7_YEAST	Q03735 saccharomyc
42	92.5	4.1	357	1	SFAI_STRFR	Q03424 streptomyc
43	92.5	4.1	409	1	ALP_TRIHA	Q03420 trichoderma
44	92.5	4.1	409	1	GUXX_BACSA	P05655 bacillus sp
45	92.5	4.1	950	1	CDAL_HUMAN	Q9513 homo sapien

ALIGNMENTS

RESULT	ID	GUXX_CELFI	STANDARD	PRT	872 AA.
AC	P50401	GUXX_CELFI			
DT	01-OCT-1996	(Rel. 34, Last sequence update)			
DT	01-OCT-1996	(Rel. 34, Last annotation update)			
DE	Exoglucanase A precursor (EC 3.2.1.91) (Exocellulohydrolase A)				
DE	(1,4-beta-cellulohydrolase A) (CBP95).				
GN	CBHA.				
OS	Cellulomonas fimi.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.				
OX	NCBI_TaxID=1708;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SRRAIN-ATCC 484;				
RX	MEDLINE=94344030; PubMed=8065260;				
RA	Meinke A., Gilkes N.R., Kwan E., Kilburn D.G., Warren R.A.J.,				
RA	Miller R.C., Jr.;				
RT	"Cellulohydrolase A (Cbha) from the cellulolytic bacterium				
RT	Cellulomonas fimi is a beta-1,4-exocellulohydrolase analogous to				
RT	Trichoderma reesei CBH II.";				
RL	Mol. Microbiol. 12:413-422(1994).				
RN	[2]				
RP	SEQUENCE OF 41-58.				
RX	MEDLINE=93209933; PubMed=8458833;				
RA	Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C., Jr., Warren R.A.J.;				
RT	"Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase				
RT	D (Cend), a family A beta-1,4-glucanase.";				
RL	J. Bacteriol. 175:1910-1918(1993).				
CC	- FUNCTION: THIS ENZYME HYDROLYSES 1,4-BETA-D-GLUCOSIDIC LINKAGES OF				
CC	CELLULOSE. WEAK ACTIVITY AGAINST CARBOXYMETHYLCELLULOSE, BACTERIAL				
CC	MICROCRYSTALLINE CELLULOSE AND BARLEY BETA-GLUCAN. HAS ALSO WEAK				
CC	ENDOGALACTANASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS WITH INVERSION				
CC	OF ANOMERIC CONFIGURATION.				
CC	- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages				
CC	in cellulose and cellotetraose, releasing cellobiose from the non-				
CC	reducing ends of the chains.				
CC	- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.				
CC	- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN				
CC	(CBD).				
CC	- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL				
CC	HYDROLASES).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; L25809; AAC36898.1; -				
DR	HSSP; P07986; 1EXG.				
DR	InterPro; IPR001919; CBD_2.				

DR InterPro: IPR003961; FN.III.
DR InterPro: IPR003962; FN.III.repeat.
DR InterPro: IPR001524; Glyco_hydro_6.
DR Pfam: PF00553; CBD_2; 1.
DR Pfam: PF00041; fn3; 3.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PRO0014; FNTYPEIIT.
DR PRINTS: PRO0733; GLHYDRASE6.
DR PRODOM: PD003733; Glyco_hydro_6; 2.
DR SMART: SM0060; FN3; 3.
DR PROSITE: PS00561; CBD_BACTERIAL; 1.
DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; signal.
FT SIGNAL 1 40
FT CHAIN 41 872
FT DOMAIN 41 477
FT DOMAIN 478 563
FT DOMAIN 573 664
FT DOMAIN 673 768
FT DOMAIN 769 872
FT ACT_SITE 188 872
FT ACT_SITE 410 410
FT DISULFID 140 202
FT DISULFID 374 428
FT DISULFID 770 869
SQ SEQUENCE 872 AA; 89300 MW; 788384078955338 CRC64;

Query Match 55.3%; Score 1244; DB 1; Length 872;
Best Local Similarity 55.7%; Pred. No. 2,6e-82;
Matches 239; Conservative 60; Mismatches 120; Indels 10; Gaps 6;

QY 3 HYDNYAGATFFVNPYMAOEVSEANOT-NATLAKKRVSTSTYAVMDRIANGVN 61
DB 44 HVDNYAGATFFVNPYMAOEVSEANOT-NATLAKKRVSTSTYAVMDRIANGVN 103
QY 62 GGGPTTYLDAALSOOGGT-TPREYIEIYVDLGRDCAALASNGELPNTAALQYETGY 120
DB 104 DONGLAKFHDNAVAOQKAGVPLVNLVLYDLPGRDCAALASNGELPNTAALQYETGY 163
QY 121 IDPIASILSNPKYSSIRIYITIEPDSLPAVYVNMISIQACATVAPYVEGIEYALTKLHAI 180
DB 164 IDPIADLDNPEYESIRIATIEPDSLRLTINISEPNOQAAPYRQGVKATLDRKLNHI 223
QY 181 PNVYTYMDAASHGWLGPNNNA---SGYVOEVOKVLINASTGVNGIDGFTVNTANVPLKEP 237
DB 224 PNVYTYMDAASHGWLGPNNNA---SGYVOEVOKVLINASTGVNGIDGFTVNTANVPLKEP 281
QY 238 FMT-ATQOYGGQPVESANFYOMNPIDIEADYAVDLYSRVYAGPSSIGMLDITLRNGWG 296
DB 282 LISDSSLINNTPIRSSKFTENFDELDYTAHMRRLVLAAGPSSIGMLDITLRNGWG 341
QY 297 GNEPTGSTATVNTFYVNSKIDLRHGRGLVNCNONGAGIGOPPOASPTDFPNHLDAYV 356
DB 342 GNRPTSTIASDVNAVYANRVRHVGANCNPLGAGIGFETEPTEGVAASHLDAYV 401
QY 357 WKPPEGSDEGTSAAADPTTGKSDPMDCPYTTTS--YGVVTLALPNSPIAGCWPAQFDQ 414
DB 402 WKPPEGSDEGTSAAADPTTGKSDPMDCPYTTTS--YGVVTLALPNSPIAGCWPAQFDQ 461
QY 415 LVANARPAV 423
DB 462 LVKNAPVPI 470

DE Exoglucanase 3 precursor (EC 3.2.1.91) (Exocellobiohydrolase 3)
DE (1,4-beta-celllobiohydrolase 3).
GN CEL3.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 255-277 AND 331-351.
RC STRAIN=D649;
RX MEDLINE=94368092; PubMed=8085821;
RA Chow C.-M., Yague E., Raguz S., Wood D.A., Thurston C.F.;
RT "The cel3 gene of Agaricus bisporus codes for a modular cellulase and
is transcriptionally regulated by the carbon source.";
RL Appl. Environ. Microbiol. 60:2779-2785(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mycelium;
RX MEDLINE=96269930; PubMed=8662210;
RA Yague E., Chow C.-M., Challen M.P., Thurston C.F.;
RT "Correlation of exons with functional domains and folding regions in
a cellulase from Agaricus bisporus.";
RL Curr. Genet. 30:56-61(1996).
CC -! FUNCTION: SHOWS ENZYMIC ACTIVITY TOWARDS CRYSTALLINE CELLULOSE.
CELLULOSE AND BARLEY B-GLUCAN.
CC -! CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
in cellulose and cellobiose, releasing cellobiose from the non-
reducing ends of the chains.
CC -! SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
HYDROLASES).
CC -! SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: L24519; AAA50607.1; -;
DR EMBL: L24520; AAA50608.1; -;
DR EMBL: Z34007; CAA83971.1; -;
DR HSSP: P00725; I426.
DR InterPro: IPR002524; CBD_fungal.
DR InterPro: IPR001524; Glyco_hydro_6.
DR Pfam: PF00734; CBD_1; 1.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PRO0733; GLHYDRASE6.
DR PRODOM: PD001821; CBD_fungal; 1.
DR PRODOM: PD003733; Glyco_hydro_6; 1.
DR SMART: SM00236; fCBD; 1.
DR PROSITE: PS00562; CBD_FUNGAL; 1.
DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 20
FT CHAIN 21 438
FT DOMAIN 21 59
FT DOMAIN 60 87
FT DOMAIN 88 438
FT ACT_SITE 215 215
FT ACT_SITE 393 393
FT DISULFID 28 45
FT DISULFID 39 55
FT DISULFID 170 229
FT DISULFID 360 407
FT VARIANT 133 133
FT VARIANT 152 152
FT VARIANT 244 244
FT VARIANT 248 248
FT VARIANT 398 398
FT VARIANT N -> S.

01-AUG-1988 (Rel. 08, Last sequence update)
 01-MAR-2002 (Rel. 41, Last annotation update)
 Exoglucanase II precursor (EC 3.2.1.91) (Exocellulohydrolase II)
 (CBHII) (1,4-beta-cellobiohydrolase).
 CBH2.
 Trichoderma reesei (Hypocrea jecorina).
 Oos Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Hypocreaceae; Hypocrea.
 NCBI_TaxID=51453;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VTT-D-80133;
 RA MEDLINE=87248061; PubMed=3596237;
 RT Teeri T.T., Lehtovaara P., Kauppinen S., Salovuori I., Knowles J.,
 RT "Homologous domains in Trichoderma reesei cellulytic enzymes: gene
 RT sequence and expression of cellobiohydrolase II.";
 RL Gene 51:43-52(1987).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OM9414 / RUT C-30;
 RA Chen C.M., Gritzall M., Stafford D.W.;
 RT "Nucleotide sequence and deduced primary structure of
 RT cellobiohydrolase II from Trichoderma reesei.";
 RL Biotechnology 5:274-278(1987).
 [3]
 RP SEQUENCE OF 25-44.
 RA Fagerstam L.G., Pettersson L.G.;
 RT "The 1,4-beta-glucan cellobiohydrolases of Trichoderma reesei OM
 RT 9414.";
 RL FEBS Lett. 119:97-100(1980).
 [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=90333255; PubMed=2377893;
 RA Rouvinen J., Bergfors T., Teeri T.T., Knowles J.K.C., Jones T.A.;
 RT "Three-dimensional structure of cellobiohydrolase II from Trichoderma
 RT reesei.";
 RL Science 249:380-386(1990).
 [5]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=97029636; PubMed=8873646;
 RA Koyula A., Reinikainen T., Ruohonen L., Valkeajarvi A.,
 RA Claessens M., Telemann O., Kleywegt G.J., Szaendings M., Rouvinen J.,
 RA Jones T.A., Teeri T.T.;
 RT "The active site of Trichoderma reesei cellobiohydrolase II: the role
 RT of tyrosine 169.";
 RL Protein Eng. 9:691-699(1996).
 -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 (2) EXOCELLULOHYDROLASES THAT CUT THE DISSACCHARIDE CELLULOSE
 FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
 SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
 -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 in cellulose and cellobiose, releasing cellobiose from the non-
 reducing ends of the chains.
 -1- SUBCELLULAR LOCATION: Secreted.
 -1- MISCELLANEOUS: T. RESEI PRODUCES TWO DIFFERENT
 EXOCELLULOHYDROLASES. THEY ARE UNIQUE IN THAT THEY CAN HYDROLYSE
 CRYSTALLINE CELLULOSE IN THE ABSENCE OF ENDOGLUCANASES.
 -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
 HYDROLASES).
 -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).

 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M16190; AAA34210.1; .

DR EMBL; M55080; AAA72922.1; .
 DR PIR; A26472; A26472.
 DR PIR; A26160; A26160.
 DR PDB; 3CBH; 15-JAN-91.
 DR PDB; 1CB2; 19-MAR-99.
 DR InterPro; IPR000254; CBD_fungal.
 DR InterPro; IPR001524; Glyco_hydro_6.
 DR Pfam; PF00734; CBD_1; 1.
 DR Pfam; PF01341; Glyco_hydro_6; 1.
 DR PRINTS; PR00733; GLHYDRASE6.
 DR ProDom; PD001821; CBD_fungal; 1.
 DR ProDom; PD003733; Glyco_hydro_6; 1.
 DR SMART; SM00236; fcbd; 1.
 DR PROSITE; PS00562; CBD_FUNGAL; 1.
 DR PROSITE; PS00555; GLYCOSYL_HYDROL_F6_1; 1.
 DR PROSITE; PS00556; GLYCOSYL_HYDROL_F6_2; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal;
 KW 3D-structure.
 FT SIGNAL 1 24
 FT CHAIN 25 471
 FT DOMAIN 25 65
 FT DOMAIN 66 106
 FT DOMAIN 107 471
 FT ACT_SITE 199 199
 FT ACT_SITE 245 245
 FT ACT_SITE 425 425
 FT DISULFID 34 51
 FT DISULFID 45 61
 FT DISULFID 200 259
 FT DISULFID 392 439
 FT CARBOHYD 38 38
 FT CARBOHYD 111 111
 FT CARBOHYD 121 121
 FT CARBOHYD 130 130
 FT CARBOHYD 133 133
 FT CARBOHYD 134 134
 FT CARBOHYD 139 139
 FT CARBOHYD 313 313
 FT CARBOHYD 334 334
 FT CARBOHYD 199 199
 FT MUTAGEN 245 245
 FT CONFLICT 359 359
 FT CONFLICT 449 449
 FT CONFLICT 471 471
 FT SEQUENCE 471 AA; 49653 MW; C471BC35BDB8 CRC64;
 SQ
 Query Match 24.3%; Score 546; DB 1; Length 471;
 Best Local Similarity 33.6%; Pred. No. 3e-32;
 Matches 145; Conservative 51; Mismatches 154; Indels 82; Gaps 13;
 OY 1 ATHVDNPYAGATFFVNPYVAQVQSEANQNTAAKRVYSTYSTAVWMD----- 52
 DB 110 ATYSGNPFVGPWMAVAYTSEVSSLAISLGAMATAAATAVAKVPSFMDLTLPML 169
 OY 53 --RIAAINVGNGPGLITLDAALSSQOCTPEVIEYIVYDIPGRCAALASNGELPATA 110
 DB 170 EQLTDIRTANKNG--NTAG-----QFVYTLDPDRCAALASNGEYSIND 213
 OY 111 AGLQETQYIDIPIASILSPYSSLRVITIEPDSLPANVMSIQACATVPYEEGI 170
 DB 214 GGVAKTK-NYIDITQIV--VEYSDIRTLIVIEPDSLANVNLGPKCANQASVLECI 270
 OY 171 EYALFKLAIPNVYLYMDAHSGLGMPNNAAGVQVQKVLNASTIGVNGIDGVTNTAN 230
 DB 271 NVAAYTQLN-LPNAVAYLDAGHGAGWPNQDPAQLFANVYKKNMSSPALGLATNVAN 329
 OY 231 YTPLEPFMTAQOQGGQVYESANFYQAMPDIDEADYAVDLYSRVLAAGFPSSIGMLDT 290
 DB 330 YN-----GWNITSPSYTQGNVAVYNEKLYIHAIGPLLANHGW-SNAFFITDQ 375
 OY 291 LKNGGNGPNEPTGPRATVDFVNGSKLDLRHGILWONGAGAGIGQPOPSPTDFPRA 350
 DB 376 GRSG-----KQPTG-----QQQMGWQCNVIGTIGRGIRPSANTGD--S 410

```

OY 351 HDAIVWIKPGESGTSNAPSDPTGKRSDDPCDTYTSYGVLMNALPNSPIGWQFPA 410
DB 411 LLDSTVWVKPGEGCGTSDSSAP-----RFDSHC-----ALPDALQPAPOGAMPQA 457
OY 411 QFDQDLVANARPA 422
DB 458 YFVOLLTFNANPS 469

RESULT 5
GUN2.THEFU STANDARD: PRT; 441 AA.
AC P26222:
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase E-2 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E-2)
DE (Cellulase E-2) (Cellulase E2).
GN CELB.
OS Thermomonospora fusca.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptosporangiales; Thermomonosporaceae;
OC Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YX:
RA MEDLINE-91258320; Pubmed-1904434;
RA Lao G., Giangas G.S., Jung E.D., Wilson D.B.;
RT "DNA sequences of three beta-1,4-endoglucanase genes from
RT Thermomonospora fusca.";
RL J. Bacteriol. 173:3397-3407(1991).
RN [2]
RP REVISIONS: SEQUENCE FROM N.A.
RC STRAIN-YX:
RA Jung E.D., Lao G., Irwin D., Barr B., Benjamin A., Wilson D.B.;
RA Submitted (MAY-1993) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 32-47.
RA Wilson D.B.;
RL "Cellulases of Thermomonospora fusca.";
RL Meth. Enzymol. 160:314-323(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 32-317.
RX MEDLINE-94002001; Pubmed-8399160;
RX Spezio M., Wilson D.B., Karpus P.A.;
RT "Crystal structure of the catalytic domain of a thermophilic
RT endoglucanase.";
RL Biochemistry 32:9906-9916(1993).
CC -1- CATALYTIC ACTIVITY: Endohydrolays of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PATHWAY: CELLULOSE DEGRADATION.
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M73321; AAC06388.1; -.
DR PIR; A42360; A42360.
DR PDB; 1TML; 31-JAN-94.
DR InterPro; IPR001919; CBD_2.
DR InterPro; IPR001524; Glyco_hydro_6.
DR Pfam; PF00553; CBD_2; 1.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR

```

DR	PRINTS:	PRO0733;	GLHYDRILASE6.	
DR	ProDom:	PD003733;	glyco_hydro.6;	1.
DR	PROSITE;	PS00561;	CB_D_BACTERIAL;	1.
DR	PROSITE;	PS00655;	GLYCOSYL_HYDROL_F6.1;	1.
DR	PROSITE;	PS00656;	GLYCOSYL_HYDROL_F6-2;	1.
KW	Cellulose degradation:		Hydrolase; Glycosidase; Signal;	3D-structure
FT	CHAIN	1	31	
FT	DOMAIN	32	441	ENDOGUCANASE E-2.
FT	DOMAIN	32	320	CATALYTIC.
FT	DOMAIN	321	340	LINER.
FT	DOMAIN	341	441	CELLULOSE-BINDING.
FT	ACT_SITE	110	110	
FT	ACT_SITE	148	148	PROTON DONOR.
FT	ACT_SITE	296	296	NUCLEOPHILE.
FT	DISULFID	111	156	
FT	DISULFID	263	298	
FT	DISULFID	346	438	POTENTIAL.
FT	STRAND	36	36	
FT	TURN	40	41	
FT	HELIX	43	50	
FT	TURN	52	53	
FT	TURN	55	56	
FT	HELIX	57	63	
FT	TURN	64	66	
FT	STRAND	69	69	
FT	STRAND	71	73	
FT	HELIX	78	95	
FT	TURN	96	96	
FT	STRAND	98	98	
FT	STRAND	100	103	
FT	TURN	109	112	
FT	HELIX	122	134	
FT	TURN	135	138	
FT	STRAND	142	145	
FT	TURN	147	148	
FT	HELIX	149	153	
FT	TURN	154	155	
FT	HELIX	158	178	
FT	TURN	180	181	
FT	STRAND	183	187	
FT	HELIX	196	205	
FT	TURN	206	207	
FT	HELIX	208	211	
FT	STRAND	214	217	
FT	TURN	219	220	
FT	HELIX	225	239	
FT	TURN	240	240	
FT	TURN	242	243	
FT	STRAND	245	249	
FT	TURN	254	255	
FT	TURN	259	260	
FT	TURN	266	267	
FT	STRAND	275	275	
FT	TURN	281	282	
FT	STRAND	283	288	
FT	TURN	292	293	
FT	STRAND	294	294	
FT	TURN	302	303	
FT	STRAND	305	305	
FT	HELIX	307	315	
FT	TURN	316	316	
Q0	SEQUENCE	441 AA;	45843 MW;	87218B4537092AE5 CRC64;

Query Match	11.6%	Score 261.5	DB 1	Length 441
Best Local Similarity	26.3%	Pred. 20.9	9.9e-12	
Matches	115	Conservative	48	Mismatches 127
				Indels 147
				Gaps 27
QY	5	DNRYGATGFEFFNP-----YAAQEVSEAAQCNATNLAAKMRVSYSTRVMDRIATAT-N	58	
Db	33	DSP-----FYANPNSSAKEMVR-----NNPNDP-----RTFYVIDRIASVYO	69	
QY	59	G---VNGSPG-LTTYLDALASQOQGTTEVEIEIVYDLPGRDCAALASNGELPATPAAGIQ	114	

```

Db 70 GTWFHHNPGQTGVDAALMSAQ-AAGKIRPLVYNNABGRCCGHHSSG-----APSHS 123
OY 115 TTEYOYIDPLASISLN-PRYSRLVITVITIEPDSLNAVNTMSIQACATAVPYEGCI--- 170
Db 124 AYRS-WIDEFAAGLKNRPAY-----IIVEPD-----LISLSSCMQHVQOEVIET 167
OY 171 -EYALTKLHA-IPNYIYMDAAHSGWLMPPNNAAGYVOEVQVNLASIGVNDIEFYNT 228
Db 168 MAYKALKAKAGSSQARITFDAGHSAMHS-PAOMASMLQO-----ADIS-NSAHGIATNT 219
OY 229 ANYTPLEKPEFMTATQOVGQPVESANFYOMNPDIADYADVAVDLVSRVLAAGPSSIGMLI 288
Db 220 SN-----YRWT--ADEVAVAKAVLS---ALIGNP-SLRAYI 248
OY 269 DTLRNGMGPNPEPTGPSTATDVNTVYNOSKIDLRHGLMCNONGAGLGQPPQASPTDFP 348
Db 249 DTSRNG-----NCPAG-----NE-----WCDPSGRALGTP---STTNNG 279
OY 349 NAHLAYVWIKRPGESDG-----TSAASPTTKGKSDPMCDPTTYS 390
Db 280 DMTDAFMTKLGEDGCIAGAGFPQOAYEMAIAAGTNPMPNPTPTPTPTTPP 339
OY 391 YGVLTNALPNSPIAGQW 407
Db 340 PGSSGACATATYTIANMW 356

```

```

RESULT 6
GUNA_CELFI STANDARD; PRT; 449 AA.
AC P07984;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
GN CENA.
OS Cellulomonas fimi.
OC Bacteria; Filumetes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococciaceae; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN RP MEDLINE-87055249; PubMed-3023193;
RA Wong W.K.R., Gerhard B., Guo Z.M., Kilburn D.G., Warren R.A.J.,
RA Miller R.C. Jr.;
RT "Characterization and structure of an endoglucanase gene cena of
RT Cellulomonas fimi.";
RL Gene 44:315-324(1986).
RN RP DOMAINS.
RX MEDLINE-90036847; PubMed-2681184;
RA Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT "Structural and functional analysis of a bacterial cellulase by
RT proteolysis.";
RL J. Biol. Chem. 264:17802-17808(1989).
RN RP DISULFIDE BONDS.
RX MEDLINE-92104156; PubMed-1761039;
RA Gilkes N.R., Claessens M., Aebersold R., Henrissat B., Meinke A.,
RA Morrison H.D., Kilburn D.G., Warren R.A.J., Miller R.C. Jr.;
RT "Structural and functional relationships in two families of beta-1,4-
RT glycanases.";
RL Eur. J. Biochem. 202:367-377(1991).
CC CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC CC (2) EXOCELLULOHRDROLASES THAT CUT THE DISACCHARIDE CELLULOSE
CC CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
CC CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic

```

```

CC CC linkages in cellulose.
CC CC -1- PTH: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
CC CC SITE FOR PROTEOLYSIS.
CC CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
CC CC HYDROLASES).
CC CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC CC (CBD).
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL: M15823; AAA23084.1;
CC CC DR PIR: A24993; A24993.
CC CC DR HSSP: P07986; 1EXG.
CC CC DR InterPro: IPR001919; CBD_2.
CC CC DR InterPro: IPR001524; glyco_hydro.6.
CC CC DR Pfam: PF00553; CBD_2; 1.
CC CC DR Pfam: PF01341; Glyco_hydro.6; 1.
CC CC DR PRINTS: PR00733; GLHYDRASE6.
CC CC DR ProDom: PD003733; Glyco_hydro.6; 1.
CC CC DR PROSITE: PS00651; CBD_BACTERIAL; 1.
CC CC DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
CC CC DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
CC CC KW Cellulose degradation; Hydrolyase; Glycosidase; Signal.
CC CC FT SIGNAL 1 31
CC CC FT CHAIN 32 449
CC CC FT DOMAIN 32 138
CC CC FT DOMAIN 139 168
CC CC FT DOMAIN 438 449
CC CC FT ACT_SITE 247 247
CC CC FT ACT_SITE 283 283
CC CC FT ACT_SITE 423 423
CC CC FT DISULFID 35 134
CC CC FT DISULFID 248 291
CC CC FT DISULFID 390 426
CC CC SQ SEQUENCE 449 AA; 46731 MW; 67FF887814B3348D CRC64;

```

```

Query Match 10.9%; Score 244.5; DB 1; Length 449;
Best Local Similarity 24.5%; Pred. No. 1.7e-10;
Matches 103; Conservative 43; Mismatches 128; Indels 147; Gaps 19;
OY 7 PYAGATFEFNPWAOEVOSEANOTNATLAAMRVY-STYSTAVWMDRTAINGVCGPG 65
Db 168 PQPTSGFYDPTTQGYRAMQASGIDKALLEKIALTPQAYWGNWADASHAQAEV----- 222
OY 66 LTTTIDALISQOQGTTEPEVIEIYVDLPGRDCALASNGELPATAGLOTOTYOTIDPIA 125
Db 223 -ADYGRAVA--AGKP---MLVVAIPGRDCGSHSGGVSEVA-----RWVDIVA 269
OY 126 -SILSNPKXSILRIYITIEPDSLPAVNTMSIQACATFANPYEGC-----EYALTKLHA 179
Db 270 QGTRGNP-----IVILEPDAL-----AQLGDS-----GQGRVGFLLAKSL-T 309
OY 180 IPNVYIYMDAAHSGWLMPPNNAAGYVOEVQVNLASIGVNDIEFYNTANTPTLKEPMP 239
Db 310 LAGARYIDAGRAKWL-----VDTPVKNLN-OVGEYAVGFRALNTSNQ-----T 354
OY 240 TATQOVGQPVESANFYOMNPDIADYADVAVDLVSRVLAAGPSSIGMLIDTLRNGMGPN 299
Db 355 TADSKAYGQGIS-----QRLGCKKF-----VIDTSRNGNS-- 385
OY 300 EPLGPESTADVNTFVNSKIDLRHGLMCNONGAGLGQPPQASPTDFPNAHLDAYWIK 359
Db 386 -----NGEWCNPGRALGEPVAVN---DGSGLDALLMWK 417
OY 360 PGESDGTSAASPTTKGKSDPMCDPTTYSGVLTNALPNSPIAGQWPPAFQDOUVANA 419
|||||

```

```

Db 418 LPEBDSG-----ACNGRPAQGMQEIALLMARNA 447
QY 420 R 420
Db 448 R 448

RESULT 7
GUNA_MICBI STANDARD; PRT; 456 AA.
ID GUNA_MICBI
AC P26414:
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 37, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
GN CELA.
OS Microbisporea bispora.
OC Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;
OC Actinomycetales; Streptosporangineae;
OC unclassified Streptosporangineae; Thermobispora.
OX NCBI_TaxID=2006;
RN [1]
RA Sequence from N.A.
RA Yablonsky M.D., Elliston K.O., Eveleigh D.E.;
RL (in) Coughlan M.P. (eds.);
RL Production, characterization and application of cellulose,
RL hemicellulose and lignin enzyme degrading systems, pp.77-83, Elsevier,
RL London (1989).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC HSSP: P26222; 1TM.
DR Interpro: IPR001919; CBD_2.
DR Interpro: IPR001524; Glyco_hydro_6.
DR Pfam: PF00553; CBD_2; 1.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PR00733; GLYHDRLASE6.
DR PRODOM: PD003733; Glyco_hydro_6; 1.
DR PROSITE: PS00361; CBD_BACTERIAL; 1.
DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
DR KEGG: Cellulose degradation; Hydrolase; Glycosidase; signal.
FT SIGNAL 1 30
FT CHAIN 31 456
FT DOMAIN 31 322
FT DOMAIN 323 355
FT DOMAIN 356 456
FT ACT_SITE 113 113
FT ACT_SITE 151 151
FT ACT_SITE 300 300
FT ACT_SITE 114 159
FT DISULFID 267 302
FT DISULFID 360 453
FT SEQUENCE 456 AA; 47011 MW; B06D8595E322848F CRC64;

Query Match 10.6%; Score 238.5; DB 1; Length 456;
Best Local Similarity 25.4%; Pred. No. 4.7e-10;
Matches 121; Conservative 41; Mismatches 125; Indels 189; Gaps 27;

QY 5 DNYVAGTFVFNPNYAGVQSEANQNTATLAKMRVSYSTYAVMMDRITAI-NG---V 60
Db 36 DSP-----FVDP-----OSNAKVVAA------PNDPRFVIRDRITAAVPTGKMPA 77
QY 61 NGSPG-----LTTYLDALASQOGTPEVEIYVYDLPGRDCAALASNGELPATPAAGIQ 115
Db 78 NYNPSYRAVDAVYGAAN--AGKIP---IMVYAMPNDCGSPSAG-----ARNHT 126
QY 116 YETQYIDPIASIIISNPKYSSLRITVTIIEPDSLPAVATNMS-----IQACATAVPYEKG 169

```

```

Db 127 AYRAMIDEIAGLRNRP-----AVIIIEPDALPTMTCMSPSEQAEOVASVGA-----G 176
QY 170 IEYALTKLHAIPNYIYMDAAHSGMLGPNNAAGVQVQEVORLNASIGVNGIDEFTNTA 229
Db 177 KFKFAASSQA-----KYFEDACHDAWVPADEMAS-----RLRADT-ANSADIALNVS 224
QY 230 NYTPLEKPEMTATQOVGGQPVESANFYQWNPDIIDEADVADVLSRLVAAGPSSIGMLID 289
Db 225 NY-----RYTSGL-----ISYAKSVLSAIG-----SHLRVID 253
QY 290 TLNRMGCGNPEPPGSPATDVNTVFVNSKIDLRHGRGLMCNAGAGCOPQASPTPEPN 349
Db 254 TSRNG-----NGPLGSE-----WCDPPGRATG---TWSTTPTGD 284
QY 350 AHDYAVYKRPGEESDTS-----ASDP-----TGKSDP 381
Db 285 PAIDAFIMIKRPGADDCIATPGVFVFDRAVELAMNAAPPTSPSPPTSPSPSOSP 344
QY 382 -----WCDPYTTSYGVLTNLP-----NSPIAG---QW 407
Db 345 GSPSPSPSPAPGRACATY-----ALVNWPGGFGQAEVTKNKGSSPIMGWTVQW 395

RESULT 8
GUNI_STRSQ STANDARD; PRT; 359 AA.
ID GUNI_STRSQ
AC P13933:
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Endoglucanase 1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase) (Carboxymethyl cellulase) (CMCase I).
GN CASA.
OS Streptomyces sp. (strain KSM-9).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1931;
RN [1]
RA Sequence from N.A., AND SEQUENCE OF 71-95.
RA Nakai R., Horiouchi S., Beppu T.;
RT "Cloning and nucleotide sequence of a cellulase gene, casa, from an
RT alkalophilic Streptomyces strain.";
RL Gene 65:229-238(1988).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RX MEDLINE-93138417; Pubmed-8422992;
RA Damude H.G., Gilkes N.R., Kilburn D.G., Miller R.C. Jr.;
RA Warren R.A.;
RT "Endoglucanase Casa from alkalophilic Streptomyces strain KSM-9 is a
RT typical member of family B of beta-1,4-glucanases.";
RL Gene 123:105-107(1993).
RN [3]
RP SEQUENCE OF 185-222, AND REVISIONS.
RX MEDLINE-92104156; Pubmed-1761039;
RA Gilkes N.R., Claessens M., Aebersold R., Henrisat B., Melnke A.,
RA Morrison H.D., Kilburn D.G., Warren R.A.J., Miller R.C. Jr.;
RT "Structural and functional relationships in two families of beta-1,4-
RT glycanases.";
RL Eur. J. Biochem. 202:367-377(1991).
CC -1- FUNCTION: CMCase I PREFERENTIALLY HYDROLYZES CARBOXYMETHYL
CC CELLULOSE (CMC) AND SHOWS A MAXIMUM ACTIVITY AT PH 8.5.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


QUTX_XANCU STANDARD; PRT; 790 AA.

AC 09XD78;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Probable quinone dehydrogenase [Pyroloquinoline-quinone] (EC 1.1.99.25).

GN QUMA.

OS Xanthomonas campestris (pv. juglandis).

OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;

OC Xanthomonas

OX NCBI_TaxID=44291;

RN 11

RP SEQUENCE FROM N.A.

RC STRAIN=C5;

RX MEDLINE=20063481; PubMed=10594704;

RA Lee Y.-A., Lo Y.-C., Yu P.-P.,

RT "A gene involved in quinone metabolism is specific to one DNA homology group of Xanthomonas campestris."

RL J. Appl. Microbiol. 87:649-658(1999).

CC -1- CATALYTIC ACTIVITY: Quinate + pyroloquinoline-quinone = 5-dehydroquinone + reduced pyroloquinoline-quinone.

CC -1- COFACTOR: PQQ (BY SIMILARITY).

CC -1- PATHWAY: QUINIC ACID CATABOLIC PATHWAY; FIRST STEP. THIS PATHWAY ALLOWS GROWTH OF BACTERIA WITH QUINATE BY ITS CONVERSION TO PROTOCATHECHUATE AND SUBSEQUENT METABOLISM BY THE BETA-KETOADIPATE PATHWAY.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: BELONGS TO THE BACTERIAL PQQ DEHYDROGENASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AF109471; AAC38453.1; -

DR InterPro: IPR001479; Bac_PQQ.

DR InterPro: IPR002372; Bac_PQQ_repeat.

DR Pfam: PF01011; Bacterial_PQQ_6.

DR PROSITE: PS00363; BACTERIAL_PQQ_1; FALSE_NEG.

DR PROSITE: PS00364; BACTERIAL_PQQ_2; FALSE_NEG.

KW Oxidoreductase; PQQ; Quinate metabolism; Transmembrane.

FT TRANSMEM 22 42 POTENTIAL.

FT TRANSMEM 48 68 POTENTIAL.

FT TRANSMEM 77 94 POTENTIAL.

FT TRANSMEM 106 126 POTENTIAL.

SO SEQUENCE 790 AA; 82896 MW; B75F29B52A49FE6F CRC64;

Query Match 5 0%; Score 112; DB 1; Length 790;

Best Local Similarity 23.3%; Pctd. No. 1.4; Indels 144; Gaps 24;

Matches 107; Conservative 39; Mismatches 170;

QY 43 STYSTAVMDRIAINGVN-----GGGLTYYIDALSSQOQT---TP-EVIRIYI 89

DB 152 SATGNIDGSRFPAALDQINSNGRPAAGSPPTTGGRIANSDGAGADQLTPLOYGEKVF 211

QY 90 YDLPGRDCAAL-ASNG-----ELPATVA-----GLQTYETQYIDYDIASLSPKYSYL 136

DB 212 LCTPHNNLILALDASTGKOLMRREINATTSVMQRCRGLGFDDADALPAPSV-ANP--SPI 268

QY 137 RYITIIERDLPNAVNTMSIQACATAPRYEGEIVALTMLHALPNYIYIMDAHSGMLG 196

DB 269 AATYVAGACRCRRRLFTNTIDGRILAV-----DAQTGA-- 302

QY 197 WPNNASGYVOEVQKVLNASIGVNGIDGFVNTANTYPLKEPFMTATQOVGQ----- 248

DB 303 -----CGFGSNGVDLKAIGLA-APDPF-----YQLTSPVLVAGTIVGGRTRADNDVQ 351

QY 249 -----PVSAFYQNNPDIDEADYAVDLKSLVA-ACPPSSIGMLIDTLRN 293

DB 352 TDMPGVYRGSMSPVRSAGIDPCNPHDRQAPAGSSVSRSTPVWMAPMSDAAMTVFL 411

QY 294 GMGSPN-----EPT-----GPST-ATDVNT-----EVNOSKIDLRHGRIGMCNNGAGL 336

DB 412 PLGGPSYDLGAEKRTALDHRYGASVIALDATGTGAEKRWYOT-----VHNDLM-----DF 460

QY 337 GQPPQASPTDFPN---AHLDAYVWIKPPGESDGTSAASDPPTGKSPMCD----- 384

DB 461 DLPMPQSLIDEPNDGSHTRAV-----IGTKAGIYIVLRATGR---PLTEVREVPYKGS 513

QY 385 -----PTYTSYGVLTNALPNSPIAGWQFPAQFDQLV 416

DB 514 DIAHQYAPTOPLSVGMPOIGTKHLTESDMWGATAMDQML 553

RESULT 11

CN16_HAEIN STANDARD; PRT; 657 AA.

AC P44764;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE 2',3'-cyclic-nucleotide 2'-phosphodiesterase precursor (EC 3.1.4.16).

GN CPDB OR H10583.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus.

OX NCBI_TaxID=727;

RN 11

RP SEQUENCE FROM N.A.

RT STRAIN=RD / KW20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,

RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,

RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,

RA Fine L.D., Frichman J.L., Fuhrmann J.L., Geoghagen N.S.M.,

RA Greth C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,

RA Venter J.C.;

RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";

RL Science 269:496-512(1995).

CC -1- FUNCTION: THIS BIFUNCTIONAL ENZYME CATALYZES TWO CONSECUTIVE REACTIONS CONVERTING 2',3'-CYCLIC-NUCLEOTIDE TO 3'-NUCLEOTIDE AND THEN 3'-NUCLEOTIDE TO NUCLEIC ACID AND PHOSPHATE (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: Nucleoside 2',3'-cyclic phosphate + H(2)O = nucleoside 3'-phosphate.

CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).

CC -1- SIMILARITY: BELONGS TO THE 5'-NUCLEOTIDASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: U32740; AAC22242.1; -

DR TIGR: H10583; -

DR InterPro: IPR002224; 5_nucleotidase.

DR InterPro: IPR000934; Ser_cyt_p_hosphatase.

DR Pfam: PF01009; 5_nucleotidase; 1.

DR Pfam: PF02872; 5_nucleotidase; 1.

DR PROSITE: PS00785; 5_NUCLEOTIDASE_1; 1.

DR PROSITE: PS00786; 5_NUCLEOTIDASE_2; 1.

KW Hydrolyase; Multifunctional enzyme; Periplasmic; Signal; Complete proteome.

FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 657 2',3'-CYCLIC-NUCLEOTIDE 2'-
 FT SEQUENCE 657 AA; 72763 MM; 201CAB415014499 CRC64;
 SO
 Query Match 4.8%; Score 107; DB 1; Length 657;
 Best Local Similarity 21.0%; Pred. No. 2.5;
 Matches 75; Conservative 33; Mismatches 123; Indels 126; Gaps 16;

QY 166 YEOGIEYALTKLHAIPNYIYDA--AHSGWL-----GMPNNAS 202
 DB 243 YOEGBENAFYADVPHT---DAVIFGSHSLFPNKEFASPNADIYNGITKGIPESEA 298
 QY 203 GT-----VOEVQVYNASISVNGIDGVTNTANTYPIKEPFMTAT 242
 DB 299 GYWMANNISVVDLGLTEHGKMIYTGKAVLRPIYDIETKKALAKNDPETALLKPVHEAT 358
 QY 243 QOVGQPYESA--NFYQ-----NPDIADYAVDLSRLVAAGFPSSIGMLITLRNG 294
 DB 359 KRYVSOPYGRKADNMYSTIALLDQDPTIYVQAQKAYEKVAPSIAAMAGLPI--LSAG 416
 QY 295 -----WGGPNEPTGPSTADVTNFEVNSKIDLRHRLMCNQN-----G 333
 DB 417 APFKAGGRKNDPTG-----YTEVKNKGLTFRNADLVLYPNLTVVKATGEQLKEWLE 469
 QY 334 AGLGPPQASPT-DEPNMHLDAIYV-----IKPGESDGTSAASDP 373
 DB 470 CSAGMEKQIDPDSKPOSIDID--WEGFRYVFDVIDGVNVEYDLTKPARYDGECKLINP 526
 QY 374 TT-----GKSDPDCDPYTTSGYGLTNALPNSPIAGOWPFAQDQIVANARP 421
 DB 527 ESHRYVNLTYGCKPVDPRAE-----FLIATN--NYRAYGNKPGTGDKHIVYASP 574

RESULT 12
 ID GAQA_DACDE STANDARD: PRT: 680 AA.
 AC 001745;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Galactose oxidase precursor (EC 1.1.3.9) (GAO).
 GN GAO.
 OS Dactylium dendroides (Cladobotryum dendroides).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Hypocreaceae; Hyphomycetes.
 OX NCBI_TaxID=5132;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=92235025; PubMed=1569070;
 RA McPherson M.J., Ogel Z.B., Stevens C.E., Yadav K.D.S., Keen J.N.,
 RT Knowles P.F.;
 RT "Galactose oxidase of Dactylium dendroides. Gene cloning and sequence
 RT analysis";
 RL J. Biol. Chem. 267:8146-8152(1992).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
 RX MEDLINE=91163641; PubMed=2002850;
 RA Ito N., Phillips S.E.V., Stevens C.E., Ogel Z.B., McPherson M.J.,
 RA Keen J.N., Yadav K.D.S., Knowles P.F.;
 RT "Novel thioether bond revealed by a 1.7 A crystal structure of
 RT galactose oxidase";
 RL Nature 350:87-90(1991).
 CC -1- FUNCTION: CATALYSES THE STEROSPECIFIC OXIDATION OF A BROAD RANGE
 CC OF PRIMARY ALCOHOL SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: D-galactose + O(2) -> D-galacto-hexodialose +
 CC H(2)O(2).
 CC -1- COFACTOR: COPPER.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: CONTAINS 3 KELCH REPEATS.
 CC -1- DATABASE: NAME=Worthington enzyme manual;

CC WWW="http://www.worthington-biochem.com/manual/G/GAO.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M86819; AAL16228.1; -.
 DR PDB: 1GOF; 31-JAN-94.
 DR PDB: 1GOG; 31-JAN-94.
 DR PDB: 1GOH; 31-JAN-94.
 DR InterPro: IPR000421; FA58_C.
 DR InterPro: IPR001798; Kelch.
 DR Pfam: PF01344; Kelch. 3.
 DR PRINTS: PR00501; KELCHREPEAT.
 DR SMART; SM00231; FA58C; 1.
 KW Oxidoreductase; Signal; Copper; Repeat; 3D-structure.
 FT SIGNAL 1 41
 FT CHAIN 42 680 GALACTOSE OXIDASE.
 FT REPEAT 279 321 KELCH 1.
 FT REPEAT 323 372 KELCH 2.
 FT REPEAT 492 544 KELCH 3.
 FT DISULFID 59 68
 FT THIOETH 269 313
 FT DISULFID 556 559
 FT METAL 313 313
 FT METAL 536 536 COPPER.
 FT METAL 537 537 COPPER.
 FT METAL 622 622 COPPER.
 FT STRAND 48 49
 FT STRAND 56 59
 FT STRAND 63 63
 FT STRAND 64 65
 FT STRAND 66 66
 FT HELIX 68 72
 FT TURN 76 77
 FT STRAND 80 81
 FT HELIX 85 87
 FT STRAND 94 112
 FT TURN 120 121
 FT STRAND 123 123
 FT STRAND 125 131
 FT STRAND 142 145
 FT STRAND 153 169
 FT STRAND 172 172
 FT TURN 173 174
 FT STRAND 180 181
 FT STRAND 183 188
 FT TURN 197 198
 FT STRAND 201 207
 FT STRAND 214 217
 FT TURN 219 221
 FT STRAND 224 227
 FT TURN 236 237
 FT STRAND 243 249
 FT TURN 250 253
 FT STRAND 254 260
 FT TURN 262 263
 FT STRAND 271 274
 FT TURN 276 277
 FT STRAND 279 283
 FT TURN 285 286
 FT TURN 288 289
 FT STRAND 291 295
 FT TURN 296 299
 FT STRAND 300 303
 FT STRAND 312 312
 FT STRAND 315 318
 FT TURN 320 321
 FT STRAND 324 327

FT	STRAND	330	330
FT	STRAND	332	332
FT	STRAND	340	340
FT	TURN	345	348
FT	STRAND	349	353
FT	TURN	354	355
FT	HELIX	359	361
FT	TURN	366	368
FT	HELIX	369	372
FT	TURN	373	373
FT	STRAND	378	380
FT	STRAND	382	384
FT	STRAND	386	388
FT	STRAND	394	399
FT	TURN	402	403
FT	STRAND	405	411
FT	STRAND	413	414
FT	TURN	415	416
FT	STRAND	417	418
FT	STRAND	423	423
FT	TURN	424	424
FT	STRAND	426	431
FT	STRAND	432	435
FT	TURN	436	440
FT	STRAND	443	443
FT	STRAND	450	450
FT	STRAND	452	452
FT	STRAND	455	459
FT	TURN	463	464
FT	STRAND	468	471
FT	TURN	473	474
FT	STRAND	480	481
FT	STRAND	484	487
FT	TURN	489	490
FT	STRAND	493	496
FT	STRAND	499	499
FT	STRAND	501	501
FT	TURN	504	505
FT	STRAND	510	510
FT	STRAND	515	518
FT	TURN	519	522
FT	STRAND	523	526
FT	TURN	536	537
FT	STRAND	538	542
FT	TURN	544	545
FT	STRAND	548	552
FT	TURN	557	558
FT	STRAND	555	570
FT	HELIX	572	574
FT	STRAND	575	575
FT	TURN	577	578
FT	STRAND	581	581
FT	STRAND	587	590
FT	STRAND	594	596
FT	TURN	597	598
FT	STRAND	600	605
FT	STRAND	611	616
FT	STRAND	619	621
FT	TURN	622	623
FT	STRAND	624	624
FT	STRAND	630	633
FT	STRAND	636	637
FT	STRAND	642	646
FT	TURN	651	653
FT	STRAND	657	664
FT	TURN	666	667
FT	STRAND	670	670
FT	STRAND	674	679
SQ	SEQUENCE	680 AA: 72823 MW: 2F97C561B63E46E9 CRC64;	

Query Match 4.8%; Score 107; DB 1: Length 680;
Best Local Similarity 21.4%; Pred. No. 2.6;

Matches 109; Conservative 48; Mismatches 156; Indels 196; Gaps 31;			
QY	19	MAOEVS-EEANQTNATLAKMRVSTYSPAVMDRIAINGVGP-----	64
Db	55	MAVTCDSAGSGNRCN-----KAIDGNKDTFM-----HFTYGANGDPKPHPTTIDMKTT	103
QY	65	-----GLTYIDALSOQQGTTPVEIVYIDLPGRCCALASNGELPATNAGLQYETQY	120
Db	104	QNVNGL-----SMLPRDGNQNGMI-----GRHEVYLSDG-----TNW	137
QY	121	IDPIAS-----ILSNPKYSS-----LRVTITIEDSLP-NAVTKMSI-QACATAVPY	166
Db	138	GSPVAGSWPEADSTTYKYSNETRPARVRLVALTEANGQPTISIAETNWFQASSYTAPOP	197
QY	167	EOGIEYALFKLHAIPNVIYIMDA--HSG-----WLGMPNAGYQVQKVLNASIGVN	219
Db	198	GLGRMGPTIDLPVPA--AAIEPTSGRVLMSSRYRDAG-----	236
QY	220	GIDGFEVNTANTYTP-----LKEPPTATQ-----VGQPVESANFY--	256
Db	237	GSPGILITSSMDPSTGIYSDRTVTWKHDMFCPSIMDNGQIVVTGNDAKKTSLYDS	296
QY	257	---QWNPDIDEADYAVDLYRLVAGFPSSIGM---LIDPLRGM-GRPNHEPTG---PS	305
Db	297	SSDSWIPGPD-----MQYARGQSSATMSDGRVFTIGSSWSGVEKKNGEYISPS	346
QY	306	TAT-----DVNTFVNQSKIDLRHGLMCNONGAGL-----GQPPQASPTDPNHL	353
Db	347	SKTWISLPNAKVPMPLTADK-----QGLYRSDNHMLFQWKKGSVFQAQPSRAMN----	396
QY	354	AYVMIRPPGESDGTSAAS-DPTTGKRSDDPMC-----DPTY-----TT	389
Db	397	---WYTTSGSGDVKSAGKQNSNRGVAPDAMCGNVMYDAVKKILFRGSGSPDQODSDATP	453
QY	390	SYGVLT-----NALPNSPIA--GQMPPAQF	412
Db	454	NAHITLGEPTGSPNTVFASNGLYFARTF	482
RESULT 13			
ID	VG12_BP74	STANDARD:	PRT; 527 AA.
AC	P10930;		
DT	01-JUL-1989 (Rel. 11, Created)		
DT	01-JAN-1990 (Rel. 13, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Short tail fiber protein (Protein Gp12) (p12).		
GN	12.		
OS	Bacteriophage T4.		
OC	Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;		
OC	T4-like phages.		
OX	NCBI_TaxID-10665;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-D;		
RX	MEDLINE-88189824; PubMed-3357780;		
RA	Seliyanov N.A., Piliipov A.G., Mesyanzhinov V.V.;		
RA	"Nucleotide and deduced amino acid sequence of bacteriophage T4 gene		
RT	12.";		
RL	Nucleic Acids Res. 16:2334-2334(1988).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Kuter E., Arisaka F., Kunisawa T., Taugita A., Mosig G.,		
RA	Mesyanzhinov V., Ruger W., Stiham T., Thomas E.;		
RT	"Bacteriophage T4 genome analysis."		
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE OF 1-57 FROM N.A.		
RC	STRAIN-D;		
RX	MEDLINE-89356257; PubMed-2548819;		
RA	Barrett B.K., Berget P.B.;		
RT	"Using transposon Tn5 insertions to sequence bacteriophage T4 gene		
RL	11.";		

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:37:43 ; Search time 125.36 Seconds
(without alignments)
583.734 Million cell updates/sec

Title: US-09-917-384-4
Perfect score: 2249
Sequence: 1 ATWVNDNYPAGATFFVNPYWA.....AGQWPPAOFDOLVANAPAV 423

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 562222 segs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1171	52.1	683	16	09PDW2
2	1050	46.7	579	2	086730
3	999	44.4	454	2	09KH72
4	999	44.4	536	2	060029
5	704	31.3	490	3	096V97
6	692	30.8	491	3	096V98
7	690.5	30.7	351	2	09S3V4
8	618.5	27.5	460	3	002321
9	597.5	26.6	444	3	096V02
10	587.5	26.1	443	3	09CIR4
11	574.5	25.5	436	3	09P8N1
12	564	25.1	465	3	096R74
13	562.5	25.0	476	3	09CIS9
14	552	24.5	471	3	09HEX8
15	539	24.0	457	3	093837
16	491.5	21.9	472	3	093860

17	452	20.1	458	3	09UW11	09uW11 pitomyces r
18	443	19.7	460	3	09UW10	09uW10 pitomyces r
19	439.5	19.5	449	3	P78721	P78721 orpomyces r
20	436	19.4	428	3	012646	012646 neocallimas
21	417	18.5	376	3	09P808	09P808 pitomyces r
22	411	18.3	459	3	P78720	P78720 orpomyces r
23	330.5	14.7	380	16	053607	053607 mycobacteri
24	281.5	12.5	473	2	09ROE6	09ROE6 cellulomona
25	272	12.1	332	2	09RDB3	09RDB3 streptomyce
26	257	11.4	470	3	09Y894	09Y894 volvariella
27	251	11.2	383	2	09L108	09L108 streptomyce
28	244.5	10.9	457	2	053488	053488 micromonosp
29	208.5	9.3	487	2	050901	050901 myxococcus
30	120.5	5.4	1790	5	09V0H6	09V0H6 drosophila
31	120	5.3	1180	16	09CHH4	09CHH4 lactococcus
32	118	5.2	1838	5	0960C3	0960C3 drosophila
33	116.5	5.2	151	2	053652	053652 stigmatella
34	113.5	5.0	693	17	09V7F7	09V7F7 sulfolobus
35	113	5.0	638	16	09V7H10	09V7H10 clostridium
36	112	5.0	329	16	09POC8	09POC8 ureaplasma
37	111.5	5.0	391	9	09ZXR4	09ZXR4 bacterioph
38	111.5	5.0	1131	3	074851	074851 schistosach
39	111	4.9	715	10	09SEF9	09SEF9 nicotiana t
40	109.5	4.9	575	2	034938	034938 campylobact
41	109	4.8	701	2	09KW03	09KW03 pseudomonas
42	108.5	4.8	745	17	09Y9L3	09Y9L3 aeropyrum p
43	107.5	4.8	513	2	046334	046334 comamonas t
44	106	4.7	492	2	052155	052155 burkholderi
45	105.5	4.7	1217	16	097GM3	097GM3 clostridium

ALIGNMENTS

RESULT	ID	Q9PDW2	PRELIMINARY;	PRT;	663 AA.
AC	Q9PDW2	Q9PDW2			
DT	01-OCR-2000 (TREMBLrel. 15, Created)				
DT	01-OCR-2000 (TREMBLrel. 15, Last sequence update)				
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
DE	1.4-BETA-CELLULOBIOSIDASE.				
GN	XPI267.				
OS	Xyella fastidiosa.				
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;				
OC	Xyella.				
OX	NCBI_Taxid-2371;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-9A5C;				
RX	MEDLINE-20365717; PubMed-10910347;				
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,				
RA	Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,				
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,				
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,				
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,				
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Porty H.,				
RA	Faciocani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,				
RA	Fraga J.S., Franca S.C., Franco M.C., Fromme M., Furian L.R.,				
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,				
RA	Ho P.L., Hohnsels J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,				
RA	Kriegel J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,				
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,				
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,				
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,				
RA	Merck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorino C.B.,				
RA	Moore D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,				
RA	Nhani A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,				
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,				
RA	Peloso B.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,				
RA	Quaglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,				
RA	de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,				
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,				

OC Thermobifida.
 OX NCBI_TaxID=2021;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Al Y.-C., Wilson D.B.;
 RT "Genomic DNA sequence encoding Thermomonospora fusca cellulase with
 overexpression";
 RT Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
 RL EMBL: AF268074; AAF75786.1; -
 DR HSSP; P07987; 1CB2; glyco_hydro.6.
 DR InterPro: IPR001524; glyco_hydro.6.
 DR Pfam; PF01341; GLYCO_HYDRO_6; 1.
 DR PRINTS; PR00733; GLHYDRASE6.
 DR ProDom; PD003733; glyco_hydro.6; 2.
 DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
 DR SIGNAL; 1
 KW SIGNAL; 31
 FT CHAIN 32 454
 FT SIGNAL 31
 SQ SEQUENCE 454 AA; 48624 MW; F8CD1C6DA807134 CRC64;

Query Match

44.4%; Score 999; DB 2; Length 454;
 Best Local Similarity 48.3%; Pred. No. 2,2e-63;
 Matches 209; Conservative 55; Mismatches 137; Indels 32; Gaps 9;

OY 4 VDNPYAGATFFVNPYMAQVQSEANQNTAAKMRVSTYSTAVMMDRIALING---- 59
 DB 37 VNPEFGAKLYVNPWMSAKAAEPGGS-----AVANESPAVWLDRIAGIEGNDSP 86
 OY 60 VNGSGGLTYYLDAALSOQGGTTPEVIEIYIDLPGRCALASNGELPATAGLQTYETQ 119
 DB 87 TTSGMGLRHLLEAV-RQSGGDPITQVIVYINLPGRCDAALASNGELPDE--LDRIKSE 143
 OY 120 YIDPIASILSN-PKYSLSRTVITIEPDSLPAVNTMSIOACATAVPY-----YEGIEY 172
 DB 144 YIDPIDIMWDADENLRTALITIEIDSLPNTVTVNGSGTELCAVMKQNGVYNGVY 203
 OY 173 ALTKLAIPVNYIYMDAAHSGMLGMPNNSAGVQVQKVLNAS-IGVNCIDGFTVNTANY 231
 DB 204 ALTKLGEIPVNYIYMDAAHSGMLGMPNNSAGVQVQKVLNAS-IGVNCIDGFTVNTANY 263
 OY 232 TPKEPEMTATQOYGGQPVESANFYQWNPDIADYADVLSRLVYAGFPSSIGMLIDTL 291
 DB 264 SATVEPPLDVNGTVNGQLIKOSKRWDMNOYVDLSFVQDLRQALIAKGRSDIGMLIDTS 323
 OY 292 RNGMGSPNEPTGPSTATDVNTFYVQSKIDIRHGRGLMCNONGAGIGOPQASPTDFPNH 351
 DB 324 RNGMGSPNEPTGPSSSTDLNTYVDESRIHGRHPCNMNCGAGIGLGRPTVNPVNA----PG 379
 OY 352 LDAYVWIKRPGESDCTSAASPTTGKSKSDPMPCDPTVYTSYGLVLTN--ALPNSPIAGQWF 408
 DB 380 VDAYVWVWRPGESDGAISEIPNDEKGFDRMCDPTTYGNNRNGNPSGALPNAPISGHWF 439
 OY 409 PAQFDOLVANARP 421
 DB 440 SAQFRELLANARP 452

RESULT 4

PRELIMINARY; PRT; 596 AA.

060029 ID 060029
 AC 060029; PRELIMINARY; PRT; 596 AA.
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE BETA-1,4-EXOCYTOGLUCANASE PRECURSOR (EC 3.2.1.91).
 GN E3.
 OS Thermomonospora fusca.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptosporangineae; Thermomonosporaceae;
 OC Thermobifida.
 OX NCBI_TaxID=2021;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-YX;
 RA Zhang S., Lao G., Wilson D.B.;
 RT "Characterization of a Thermomonospora fusca exocellulase";
 RL Biochemistry 34:0-(1995).
 DR EMBL; U18978; AAA62211.1; -
 DR HSSP; P07986; 1EXG.
 DR InterPro: IPR001919; CBD-2.
 DR InterPro: IPR001524; glyco_hydro.6.
 DR Pfam; PF00553; CBD-2; 1.
 DR Pfam; PF01341; glyco_hydro.6; 1.
 DR PRINTS; PR00733; GLHYDRASE6.
 DR ProDom; PD003733; glyco_hydro.6; 2.
 DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
 DR SIGNAL; 1
 KW SIGNAL; 38
 FT CHAIN 39 596
 FT SIGNAL 38
 SQ SEQUENCE 596 AA; 63547 MW; B0FA527FE7721E0 CRC64;

Query Match

44.4%; Score 999; DB 2; Length 596;
 Best Local Similarity 48.3%; Pred. No. 3,2e-63;
 Matches 209; Conservative 55; Mismatches 137; Indels 32; Gaps 9;

OY 4 VDNPYAGATFFVNPYMAQVQSEANQNTAAKMRVSTYSTAVMMDRIALING---- 59
 DB 179 VNPEFGAKLYVNPWMSAKAAEPGGS-----AVANESPAVWLDRIAGIEGNDSP 228
 OY 60 VNGSGGLTYYLDAALSOQGGTTPEVIEIYIDLPGRCALASNGELPATAGLQTYETQ 119
 DB 229 TTSGMGLRHLLEAV-RQSGGDPITQVIVYINLPGRCDAALASNGELPDE--LDRIKSE 285
 OY 120 YIDPIASILSN-PKYSLSRTVITIEPDSLPAVNTMSIOACATAVPY-----YEGIEY 172
 DB 286 YIDPIDIMWDADENLRTALITIEIDSLPNTVTVNGSGTELCAVMKQNGVYNGVY 345
 OY 173 ALTKLAIPVNYIYMDAAHSGMLGMPNNSAGVQVQKVLNAS-IGVNCIDGFTVNTANY 231
 DB 346 ALTKLGEIPVNYIYMDAAHSGMLGMPNNSAGVQVQKVLNAS-IGVNCIDGFTVNTANY 405
 OY 232 TPKEPEMTATQOYGGQPVESANFYQWNPDIADYADVLSRLVYAGFPSSIGMLIDTL 291
 DB 406 SATVEPPLDVNGTVNGQLIKOSKRWDMNOYVDLSFVQDLRQALIAKGRSDIGMLIDTS 465
 OY 292 RNGMGSPNEPTGPSTATDVNTFYVQSKIDIRHGRGLMCNONGAGIGOPQASPTDFPNH 351
 DB 466 RNGMGSPNEPTGPSSSTDLNTYVDESRIHGRHPCNMNCGAGIGLGRPTVNPVNA----PG 521
 OY 352 LDAYVWIKRPGESDCTSAASPTTGKSKSDPMPCDPTVYTSYGLVLTN--ALPNSPIAGQWF 408
 DB 522 VDAYVWVWRPGESDGAISEIPNDEKGFDRMCDPTTYGNNRNGNPSGALPNAPISGHWF 581
 OY 409 PAQFDOLVANARP 421
 DB 582 SAQFRELLANARP 594

RESULT 5

PRELIMINARY; PRT; 490 AA.

096V97 ID 096V97
 AC 096V97; PRELIMINARY; PRT; 490 AA.
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE CELLOBIOHYDROLASE II-LIKE CELLULASE CELL.
 GN CELL.
 OS Orpiniomyces sp. PC-2.
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimastixales;
 OC Neocallimastixaceae; Orpiniomyces.
 OX NCBI_TaxID=50059;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PC-2;

RX MEDLINE=21405733; PubMed=11514516;
 RA Stenbäckers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
 RA Ljungdahl L.G., Op Den Camp H.J.M.; Noncatalytic docking domains of cellulosesomes of anaerobic fungi.";
 RL J. Bacteriol. 183:5325-5333(2001).
 DR EMBL; AF177205; AAL01212.1; -.
 KM Hydrolyase.
 SQ SEQUENCE 490 AA; 54051 MW; 85F40D1280759886 CRC64;

Query Match 31.3%; Score 704; DB 3; Length 490;
 Best Local Similarity 39.2%; Pred. No. 2.9e-42;
 Matches 166; Conservative 54; Mismatches 143; Indels 60; Gaps 12;

QY 2 THVDNRYAGATFEVNPYWA-OEVOSEANOTNATLAAMRVYSTSTAVMMDRIAINGVNG 61
 DB 123 THSGNFEFSVEFFLNPFYVAEVDAAIEOMTNSLKAKAEKMKYTSNAIWDITK-----N 177
 QY 62 GSGPLTLYDAALSOOGGTTPEYIEI-VIYDLPGRDCALASNGELPATAGLOTETQY 120
 DB 178 MHWLETNKAGLQOQEFEGSKVLYFVYDLPGRDCHALASNGELANDSDAQRYKTEY 237
 QY 121 IDPLASILSNPKYSLRIYITIEPDSLPAVATNM-SIQACATAPYEEQIEYALTKLHA 179
 DB 238 IDVIEEKLK--YKKSQPVVLIPEPDSLAMLTNLDSTPACRDEKYYLDGHAYLIQKIGY 295
 QY 180 INNYIYMAASGWLGMNPNASGYVOEOKVLNASIGVNG-IDGFVTYNTANTPLKEPP 238
 DB 296 LPHVAATLIDIGHAFWLGMDDNREKAGYSKYI--SSGPKGVKGTIDVNAVNTTPWEDPT 353
 QY 239 KWATQOVGQPVESANFYQWNPDIIDEADYAVDLSRLVAAGFPSSIGMLIDTLRNGMGGR 298
 DB 354 LRGPEPT-----EMNPCDEKRYLELMHKDFAAAGIESYFVCDTSRNG----- 398
 QY 299 NEPTGPSTATDVNTFVNSKIDLRHRLGIMCNGAGLQPPQASPTDPNNAHLDAYVVI 358
 DB 399 -----HKVD-RKHPEMCNQTGVGVARQDAPASVSGMD-YLDAFYVI 438
 QY 359 KPPGESDGTSAASDPTTGKSPMCDPTTTSYGLTNALPNSPIAGOWPFAQFDOLVAN 418
 DB 439 KPLGESDGT-----DTSARIDYGCGH-----DTAMKPAPEAGWFOKHFEQLENN 485
 QY 419 ARP 421
 DB 486 ANP 488
 RESULT 6
 ID 096V98 PRELIMINARY; PRT; 491 AA.
 AC 096V98:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CELLOBIOHYDROLASE II-LIKE CELLULASE CELH.
 GN CELH.
 OS Orpiniomyces sp. PC-2.
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
 OC Neocallimastacaceae; Orpiniomyces.
 OC NCBI_TaxID=50059;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PC-2;
 RX MEDLINE=21405733; PubMed=11514516;
 RA Stenbäckers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
 RA Ljungdahl L.G., Op Den Camp H.J.M.; Noncatalytic docking domains of cellulosesomes of anaerobic fungi.";
 RL J. Bacteriol. 183:5325-5333(2001).
 DR EMBL; AF177204; AAL01211.1; -.
 KM Hydrolyase.
 SQ SEQUENCE 491 AA; 53956 MW; 7C7ACCC9BFF946C4 CRC64;

Query Match 30.8%; Score 692; DB 3; Length 491;
 Best Local Similarity 39.0%; Pred. No. 2.1e-41;
 Matches 163; Conservative 53; Mismatches 142; Indels 60; Gaps 12;

QY 6 NPACATFEVNPYWA-OEVOSEANOTNATLAAMRVYSTSTAVMMDRIAINGVNG 65
 DB 130 NPFSGVFEFLNPFYVAEVDAAIOMNSLSLKAKAEKMKYTSNAIWDITK-----NMQOM 184
 QY 66 LTTYLDAALSOOGGTTPEYIEIYIDLPGRDCAALASNGELPATAGLOTETQYIDPIA 125
 DB 185 LETNLKAGALAQCGSKKVLTFVYDLPGRDCHALASNGELANDSDAQRYKTEYIDVIE 244
 QY 126 SILSNPKYSLRIYITIEPDSLPAVATNM-SIQACATAPYEEQIEYALTKLHAIPNY 184
 DB 245 EKLR--YKKSQPVVLIIEPDSLAMLTNLENPACRDEKYYLDGHAYLIQKIGVLPVHA 302
 QY 185 IYMDAASGWLGMNPNASGYVOEOKVLNASIGVNG-IDGFVTYNTANTPLKEPPFATAG 243
 DB 303 MYLIDIGHAFWLGMDDNREKAGYSKYI--SSGPKGVKGTIDVNAVNTTPWEDPTLSRCP 360
 QY 244 QVGGQPVESANFYQWNPDIIDEADYAVDLSRLVAAGFPSSIGMLIDTLRNGMGGRPEPTG 303
 DB 361 DT-----EMNPCDEKRYLEAMHKDFKAAGI-SSVYFVSDTSRNG----- 399
 QY 304 PSTATDVNTFVNSKIDLRHRLGIMCNGAGLQPPQASPTDPNNAHLDAYVVIIPPE 363
 DB 400 -----HKVD-RKHPEMCNQTGVGVARQDAPASVSGMD-YLDAFYVIRPLGE 444
 QY 364 SDGTSAASDPTTGKSPMCDPTTTSYGLTNALPNSPIAGOWPFAQFDOLVANARP 421
 DB 445 SDGTS-----DTSARIDYGCGH-----ETAMKPAPEAGWFOKHFEQLENNARP 489
 RESULT 7
 ID 09S3V4 PRELIMINARY; PRT; 351 AA.
 AC 09S3V4:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE 1,4-BETA-CELLULOBIHYDROLASE A (EC 3.2.1.91) (FRAGMENT).
 GN CELA.
 OS Cellulomonas flavigena.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococccineae; Cellulomonadaceae; Cellulomonas.
 OC NCBI_TaxID=1711;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDBB531;
 RA Herrera A., Gutierrez A., Salgado L.M., Ponce-Noyola T.;
 RT "Molecular characterization of cellulases from Cellulomonas
 RT flavigena.";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF172344; AAD48493.1; -.
 DR HSSP; P07987; 1CB2.
 DR InterPro; IPR001524; Glyco_hydro-6.
 DR Pfam; PF01341; Glyco_hydro-6; 1.
 DR PRINTS; PR00733; GLHYDRLASE6.
 DR PRODOM; PD003733; Glyco_hydro-6; 1.
 DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
 DR Hydrolyase; Glycosidase.
 FT NON_TER 351
 SQ SEQUENCE 351 AA; 37615 MW; BCDE63B65DF335C2 CRC64;

Query Match 30.7%; Score 690.5; DB 2; Length 351;
 Best Local Similarity 49.5%; Pred. No. 1.7e-41;
 Matches 151; Conservative 38; Mismatches 93; Indels 23; Gaps 9;

QY 4 VDNRYAGATFEVNPYWA-OEVOSEANOTNATLAAMRVYSTSTAVMMDRIAINGVNG 62
 DB 44 VDNRYAGAVGYVNPYWAASSSVRASAKQSDASIAAKMGYAKOPTVAVMMDRISAITGNAD 103

QY 63 GPGITTYLDAALSSQOQGT-TPEVIEIYIYDLPGRDCAALASNGELPATAGLOTYETQYI 121
 DB 104 GKLKHLDAALNAVOOQAKGRLPVLNLIYVNLPRDCYALASNGELPATDGLARYOSEYI 163
 QY 122 DPLASTLSPKYSRLRIVTIIIEPDSL.PNNAVTNMSIQACATAVPYEGGIEYALTKLHAIP 181
 DB 164 DPLVALGKPEISNLIAATIEPDSLPLNLTNISESCQOAPYRGGVAYALDKLALG 223
 QY 182 NYIYMDAAHSGMLGPNNASGYQ---EVOKVLNASIGVNGIDGFTVNTANTYPLKE-- 236
 DB 224 NYNYNIDIGHSGMLGMDNSNAGPAKIFAEVAKTTNA--GPASTIGTVSDVANTPLEEPY 281
 QY 237 PEKTAATQVGGQPEVSAN-FYQWNPDIIDEAD--YAVDLY--SRLVAAGFPSSIGMLIDL 291
 DB 282 PYPERPERSGNDPIREJOYFTSGNNELSTRNRLYGATLHNVGSFGPGFPE----- 332
 QY 292 RNCWG 296
 DB 333 RRPWG 337

RESULT 8
 ID 002321 PRELIMINARY; PRT; 460 AA.
 AC 002321;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE EXCELLOHIOHYDROLASE.
 GN CBH11.
 OS Phanerochaete chrysosporium.
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Aphyllophorales; Corticiaceae; Phanerochaete.
 OX NCBI_TaxID=5306;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9511031; PubMed=7811079;
 RA Templears C.A., Birch P.R., Sims P.F., Broda P.;
 RT "Isolation, characterization, and analysis of the expression of the
 RT cbh11 gene of Phanerochaete chrysosporium."
 RL Appl. Environ. Microbiol. 60:4387-4393(1994).
 DR EMBL: S76141; AAB32942.1; -.
 DR HSSP: P00725; 2CBH.
 DR InterPro: IPR000254; CBD_fungal.
 DR InterPro: IPR001524; Glyco_hydro_6.
 DR Pfam: PF00734; CBD_1; 1
 DR Pfam: PF01341; Glyco_hydro_6; 1.
 DR PRINTS: PR00733; GLHYDRLASE6.
 DR ProDom: PD001821; CBD_fungal; 1.
 DR ProDom: PD003733; Glyco_hydro_6; 1.
 DR SMART: SM00236; fCBD; 1.
 DR PROSITE: PS00562; CBD_FUNGAL; 1.
 DR PROSITE: PS00655; GLYCOSTYL_HYDROL_F6_1; 1.
 DR PROSITE: PS00656; GLYCOSTYL_HYDROL_F6_2; 1.
 DR Hydrolase.
 SQ SEQUENCE 460 AA; 48444 MW; ABB8A40615FE56B CRC64;

Query Match 27.5%; Score 618.5; DB 3; Length 460;
 Best Local Similarity 36.8%; Pred. No. 3,4e-36;
 Matches 154; Conservative 55; Mismatches 144; Indels 65; Gaps 12;

QY 5 DNPYACATFEVNPYNAQVSEANQTNATLAKMRVSTYPAVWMDRIAINGVNGP 64
 DB 105 NNHWGFOFLSPRYANEAANKOTIDPLTSKASVANIPTFTWLDVAKI-----P 158
 QY 65 GLTTYLDAALSSQOQGT-TPEVIEIYIYDLPGRDCAALASNGELPATAGLOTYETQYI 123
 DB 159 DLCTYLASASALCKSTGKOLVOIYIDLPRDCAAKASNGEFSIANNGQANE-NYIDQ 217
 QY 124 IASILSNPKYSRLRIVTIIIEPDSL.PNNAVTNMSIQACATAVPYEGGIEYALTKLHAIP 183
 DB 218 IYAOIO--QPPDVVAVVIEPDSLNLVNLNVQKCANAKTTLACVNTLNLAV-GV 274

QY 184 IYMDAAHSGMLGPNNASGYQEVOKVLNASIGVNGIDGFTVNTANTYPLKEPMTATQ 243
 DB 275 IYMDAGHAGMLGPNNASGYQEVOKVLNASIGVNGIDGFTVNTANTYPLKEPMTATQ 326
 QY 244 QVGGQPEVSANFYQWNPDIIDEADYADLYSRLVAAGFPSSIGMLIDLRLNGWGPNEPTG 303
 DB 327 AASDPFIT-----QGNPNYIEIHYNLAPLLOQGDWAT--FYVDGGRSG----- 370
 QY 304 PSTATDVNTFVNSKIDLRHGRGLMCNONGAGQPPQASPTDEPNHLAVYWKRPGE 363
 DB 371 -----VONIRQMGDMCNKAGAGETRP--TTNGSOFIDSIYVWKRPGE 413
 QY 364 SDGTSASDPTTGKSDPMCDPTTTSYGLVTNALPNSPIAGQWFPQPDFOLVANAR 421
 DB 414 CDGTSNNSP-----RYDSTCS-----LPDAQAPAPAGTWFQAYFQTLVSAANP 458

RESULT 9
 ID 096VU2 PRELIMINARY; PRT; 444 AA.
 AC 096VU2;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CELLULOSE CEL6B.
 GN CEL6B.
 OS Lentinula edodes (Shiitake mushroom) (Lentinus edodes).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Agaricales; Tricholomataceae; Lentinula.
 OX NCBI_TaxID=5353;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=STAMETS CS-2;
 RA Lee C.C., Wong D.W.S., Robertson G.H.;
 RT "Cloning and characterization of two cellulase genes from Lentinula
 RT edodes."
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF411251; AAK95564.1; -.
 SQ SEQUENCE 444 AA; 46368 MW; A08BEDD0835FE47C5 CRC64;

Query Match 26.6%; Score 597.5; DB 3; Length 444;
 Best Local Similarity 35.3%; Pred. No. 1e-34;
 Matches 147; Conservative 60; Mismatches 143; Indels 67; Gaps 14;

QY 6 NPYACATFEVNPYNAQVSEANQTNATLAKMRVSTYPAVWMDRIAINGVNGP 65
 DB 92 NPFTGEIYILSPYANEAIAVTOISDPTTAAAKAAKANIPTFTWLDQVAKV-----PD 145
 QY 66 LTTYL-DAALSSQOQGT-TPEVIEIYIYDLPGRDCAALASNGELPATAGLOTYETQYI 124
 DB 146 LGTYLDAASAKQKSEKNNYVOIYVYDLPRDCAALASNGELPATAGLOTYETQYI 204
 QY 125 ASILSNPKYSRLRIVTIIIEPDSL.PNNAVTNMSIQACATAVPYEGGIEYALTKLHAIP 184
 DB 205 VAOIK--QYPDVAVVAVVIEPDSLNLVNLNVQKCANAOFTTIECVYVNAQQLASV-GVT 261
 QY 185 IYMDAAHSGMLGPNNASGYQEVOKVLNASIGVNGIDGFTVNTANTYPLKEPMTATQ 244
 DB 262 MYLDAGHAGMLGPNNASGYQEVOKVLNASIGVNGIDGFTVNTANTYPLKEPMTATQ 316
 QY 245 VGGQPEVSANFYQWNPDIIDEADYADLYSRLVAAGFPSSIGMLIDLRLNGWGPNEPTG 304
 DB 317 ----DPII-----QGDPIYDEMLIETALAPIL--GSFPAH--FYVDGGRSG----- 354
 QY 305 STATDVNTFVNSKIDLRHGRGLMCNONGAGLGPQASPTDEPNHLAVYWKRPGES 364
 DB 355 -----VQDIRQMGDMCNVLAGFGTQP--TTNGSSSIDSIYVWKRPGE 398
 QY 365 DGTSAASDPTTGKSDPMCDPTTTSYGLVTNALPNSPIAGQWFPQPDFOLVANAR 421
 DB 399 DGTSTNTSSP-----RYDAHCG-----LPDATPNPAGTWFQAYFETLVERANP 442

```
RESULT 10
09C1R4 PRELIMINARY; PRT: 443 AA.
AC 09C1R4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLOBIOLHYDROLASE.
GN Chit1-1.
OS Lentinula edodes (Shiitake mushroom) (Lentinus edodes).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Tricholomataceae; Lentinula.
OX NCBI_TaxID=5353;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L54;
RA Zhao J., Deane E.E., Kwan H.S.;
RT "Molecular cloning, characterization, and differential expression of a
RL cellobiohydrolase gene from the edible mushroom Lentinula edodes.";
DR EMBL; AF244369; AAK28357.1; -.
DR HSSP; P00725; 1A26.
DR InterPro; IPR001524; Glyco_hydro_6.
DR Pfam; PF00734; CBD_1; 1.
DR PRINTS; PF01341; Glyco_hydro_6; 1.
DR PRODOM; PD001821; CBD_fungal; 1.
DR PRODOM; PD003733; Glyco_hydro_6; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
DR HydroLase.
KW HydroLase.
SQ SEQUENCE 443 AA; 46290 MM; EA9B578BF79EC25A CRC64;

Query Match 26.1%; Score 587.5; DB 3; Length 443;
Best Local Similarity 35.0%; Pred. No. 5.3e-34;
Matches 149; Conservative 60; Mismatches 146; Indels 71; Gaps 15;

OY 1 ATHVNPYAGATF---FVNPMYAOEVOSEANOTNATLAAMRVSYSTAVMMDR1AA 56
DB 82 STGTTTPAAGNPFTEQIILSPYANETIAAVTQISDPTAAAKAVANIPFTWLDQYAK 141
OY 57 INGVNGPGLFTYTL-DAALSOQOGTTPVEIEIVYDLPGRDCAALASNGELPATAG1QT 115
DB 142 V-----PDLGTYLADASAKQSEKKNLYLVYVDLPDRDCAALASNGELTIANGEAN 195
OY 116 YETQYIDPIASLSPKYSRLRTVITIEPDSLPAVNTMNSIQACATAVPYEGGIEYALT 175
DB 196 YH-DIIDIIVAKIK--QYPDVHVAVVIEPDSLAVNTLVNLSVAKANACQTYTLECVYAMQ 252
OY 176 KLHAIPNYIYVMDAHSGLMGPNNASGYVOEVOKVLNASIGVNGIOGFVNTNANTPLKPEPMATQ 235
DB 233 QLSAAV-GVTMYIDAGHAGMLGMPANLSPAOLFTSLYSNAGSPSGVRLGATVAVYANAL- 310
OY 236 EEFMTATQOVGQPYESANFYQWNPDIIDEADYAVDLSRLVAAGPSSIGMLIDTLRNGM 295
DB 311 ----VAITP---DPIT-----QGPNDYDEMLYIEALAPLL--GSPPAH--FIVDGRSG- 353
OY 296 GGPNEPPTSPSTATDVNTVFNOSKIDLRQHRGLMCNONGAGLGOPQASPTDFPNAHLDAY 355
DB 354 -----VQDIRQWGMCMVNLGAGFGTQP-----TTNTGSSLIIDSI 388
OY 356 VMKIPGSDGTSASADPTTGKKSPPMCDPYTTSYGVLTNALPSPISLAGOWFPQFQDL 415
DB 389 VMKIPGSDGTSNTSSP-----RYDHAG-----LPDATPNAPEAGTWFQAFETL 435
OY 416 VANARP 421
| | |
```

```
DB 436 VEKANP 441

RESULT 11
09P8N1 PRELIMINARY; PRT: 436 AA.
AC 09P8N1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLOBIOLHYDROLASE.
OS Coriolus versicolor.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Aphyllophorales; Coriolaceae; Coriolus.
OX NCBI_TaxID=57466;
RN [1]
RP SEQUENCE FROM N.A.
RA Novo C., Clemente A., Simoes F., Mendonca D., Matos J.;
RT "Coriolus versicolor cellobiohydrolase-like cDNA sequence.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF233583; AAF35251.1; -.
DR HSSP; P00725; 2CBH.
DR InterPro; IPR001524; Glyco_hydro_6.
DR Pfam; PF00734; CBD_1; 1.
DR PRINTS; PF01341; Glyco_hydro_6; 1.
DR PRODOM; PD001821; CBD_fungal; 1.
DR PRODOM; PD003733; Glyco_hydro_6; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
DR HydroLase.
KW HydroLase.
SQ SEQUENCE 436 AA; 46170 MM; EDECC5AFBCA324D4 CRC64;

Query Match 25.5%; Score 574.5; DB 3; Length 436;
Best Local Similarity 33.5%; Pred. No. 4.4e-33;
Matches 142; Conservative 65; Mismatches 138; Indels 79; Gaps 14;

OY 6 NPYACATFVNPMYAOEVOSEANOTNATLAAMRVSYSTAVMMDR1AINGVNGPG 65
DB 83 NPFTEQIILSPYANETIAAAVATDSSLKAKASVANIPFTWLDQYAKV-----PD 136
OY 66 LTTYL-DAALSOQOGTTPVEIEIVYDLPGRDCAALASNGELPATAG1QYETQYIDPI 124
DB 137 LGTYLADASSIQTKGQQLVPIVYELPDRDCAALASNGELSIADAGENK-DYIDQI 195
OY 125 ASILSPKYSRLRTVITIEPDSLPAVNTMNSIQACATAVPYEGGIEYALTKLHAIPNY 184
DB 196 VPOIK--QFPDVHVAVVIEPDSLAVNTLVNLSVAKANG-GTYKASVYVALQOLSSSV-GVT 251
OY 185 IYMDAHSGLMGPNNASGYVOEVOKVLNASIGVNGIOGFVNTNANTPLKPEPMATQ 244
DB 252 MYMDAGHAGMLGMPANLIDQSGSEVFAMKSAFVAFRAFAIVNREYNALTAFA----- 305
OY 245 VGGQPYESANFYQWNPDIIDEADYAVDLSRLVAAGPSSIGMLIDTLRNGMGPPREPTGP 304
DB 306 --PRPIT-----QGPNDYDEFLYIDRVRLMKLSPGPAPQ--FVVOGARG----- 346
OY 305 STATDVNFTVFNOSKIDLRQHRGLMCNONGAGLGOPQASPTDFPNAHLDAYVMKIPGGS 364
DB 347 -----QGNFRQWGMCMVNLGAGFGTGP-----TTSTGNPLDIALIIVKPGGS 390
OY 365 DGTSAADP-----TTGKKSPPMCDPYTTSYGVLTNALPSPISLAGOWFPQFQDL 418
DB 391 DGTSSSSPFRIDYSTLISVARDP-----APBAGTWFQAFETLVS 431
OY 419 -ARP 421
| |
DB 432 PTRP 435
```

RESULT 12
ID 096TP4 PRELIMINARY: PRT: 465 AA.
AC 096TP4
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE CELLOBIOHYDROLASE II.
OS Pleurotus sajor-caju (Oyster mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Pleurotaceae; Pleurotus.
OX NCBI_TaxID=50053;
RN
RP
RT Cloning a cellobiohydrolase II gene from Pleurotus sajor-caju and its
RT expression in Saccharomyces cerevisiae.
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY050517; M115038.1; -
DR EMBL: AY050518; M115037.1; -
KM Hydrolase.
SO SEQUENCE 465 AA; 49271 MW; 323AB1B235E265ED CRC64;

Query Match 25.1%; Score 564; DB 3; Length 465;
Best Local Similarity 31.9%; Pred. No. 2.7e-32;
Matches 151; Conservative 56; Mismatches 142; Indels 124; Gaps 16;

OY 6 NPAG-----ATFEVNPYMAQEOVEAANO-TNATLAKKRVYSTSTAVMMORIAI 57
DB 62 NPAGDVSHVMCOITLSPYADVEAAVAASINPALAKASVAPPIPIFEDVAKV 121
OY 58 NGVNGPGTITTYDALISOOGT-TPEVITVITDLPGRCCALASNGELPATAGLOTY 116
DB 122 -----PTLGTIYDALISTOOSTGRNOLVOIYVDLPDRCCALASNGEESIANGLANT 175
OY 117 ETOYIDPIASILSN-----PKYSLRITVITIEPDSLPAVTKMSIQACA 160
DB 176 K-NYVDQIVQIARCCPLVTSATDLACSEYQIRVAVVEPDSLANNVTMLNPKCA 234
OY 161 TAVPYEEOGIEVALTKLHAIPNYITMDAASHGLGPNNASGYVQVQVLAASIGVNG 220
DB 235 GAQAAATEGVTYALQKLMNV-GVYSYVDAGHAGLGPALGPAADLFANLYTNAGSPSF 293
OY 221 IDGVTNTANTYPLKEPFMTATQVGOPVESANFYQWNPDIIDEADV----- 267
DB 294 FRLATINNAVNYLNLNP-----SPDPVTS-----NANYDEIRITINVSDFVLWTS 340
OY 268 ---AVDLYSRVAGFPSSIGMLIDT-----LRNGMGCPNEPTGPRSTATDVNPFVQ 316
DB 341 LTICIIALAPELSSRGPRAH-FIVDGRSAVVGIRGAMD----- 379
OY 317 SKIDLOHKGKMGONAGAGIGOPPOASPTDFPAHLDAYWIRPGESDGTSAASDPTTG 376
DB 380 -----WCNVNAGGTRP-----TSTGSLDAIYVWVRPGESDGT-----DTSA 421
OY 377 KRSDPKCDPTTYSYGLTNALPNSPIA-----GQWFAPOFOLVANNAPAV 423
DB 422 VRTDGHG-----LASAKKAPAPAMASVSHSSFOYFEMLVANAVPAL 465

RESULT 13
ID 09C1S9 PRELIMINARY: PRT: 476 AA.
AC 09C1S9
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE AVICELASE 2.
GN AVI2.
OS Humicola insolens.
OC Fungi; Ascomycota; mltosporic Ascomycota; Humicola.

OX NCBI_TaxID=34413;
RN
RP
RT Cloning a cellobiohydrolase II gene from Pleurotus sajor-caju and its
RT expression in Saccharomyces cerevisiae.
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB048710; BAB39154.1; -
DR HSSP: P00725; IAZ6.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR001524; Glyco_hydro_6.
DR Pfam: PF00734; CBD_1; 1.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PR00733; GLHYDRLASE6.
DR ProDom: PD001821; CBD_fungal; 1.
DR ProDom: PD003733; Glyco_hydro_6; 1.
DR SMART: SM00236; fCBD; 1.
DR PROSITE: PS00562; CBD_FUNGAL; 1.
DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
SO SEQUENCE 476 AA; 51276 MW; D2B0054F1381E653 CRC64;

Query Match 25.0%; Score 562.5; DB 3; Length 476;
Best Local Similarity 33.7%; Pred. No. 3.6e-32;
Matches 147; Conservative 54; Mismatches 146; Indels 89; Gaps 15;

OY 1 ATHDNDYACATFEVNPYMAQEOVEAANO-TNATLAKKRVYSTSTAVMMORIAI 59
DB 113 ASTNGNPEFEGQIANNVYSEVHTLAIPOTTDALAAASAAVEPFOLDNNVTVD- 171
OY 60 VNGSGPGLITTYDALISO-----OQGTPE-VIEIVITDLPGRCCALASNGELPATAGL 113
DB 172 -----TLVETLSIEIRANQGANPYAAQIVYVDLPDRCCALASNGEESIANGLANT 223
OY 114 QTEYOTIDPIASILSPKYSILRITVITIEPDSLPAVTKMSIQACATAVPYEEOGIEYA 173
DB 224 NNK-GYINIRIELIS--FSDVRTIIVIEPDSLANNVTNNVAKCGAASVRELYIA 280
OY 174 LTKLHAIPNYITMDAASHGLGPNNASGYVQVQVLAASIGVNGIDGVTNTANT- 232
DB 281 LKQLD-LPHVAMTMDACHAGLGPANIIOPAELEPAKTYEDAGPRAVRGATVANINA 339
OY 233 -----PLKEPFMTATQVGOPVESANFYQWNPDIIDEADVAVDLYSRVAGFPSSIG 285
DB 340 WSISSEPPYSP-----NPNVDEKITYEAFPLEARGFPAG-- 376
OY 286 MLIDTLNMGCGPNEPTGPRSTATDVNPFVQSKIDLOHKGKMGONAGAGIGOPPOASPT 345
DB 377 FIVDGRSG-----KQPTG-----OKENGHCNNAIGTGFGRPTANT- 413
OY 346 DFPANHDAVWIRPGESDGTSAASDPTGKSDPKCDPTTYSYGLTNALPNSPIAG 405
DB 414 --GHQYDAFVWVRPGESDGT-----DTTAPRDYHCG-----LEDALKPAPBAG 458
OY 406 QWFAPOFOLVANNAP 421
DB 459 QWFAPOFOLVANNAP 474

RESULT 14
ID 09HEY8 PRELIMINARY: PRT: 471 AA.
AC 09HEY8
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE CELLOBIOHYDROLASE II.
GN CBH1 OR CBH2.
OS Trichoderma koningii.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.

OX NCBI_TaxID=51453, 114618;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-T. reesei;
RA Zhu L., Dong Z., Yu W.;
RT "Cloning and analysis of the CBHII gene."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-T. koningi;
RA Zhu L., Dong Z., Yu W.;
RT "Cloning and sequence analysis of the cellobiohydrolase II gene (cbh2)
RT from Trichoderma koningi K."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF302657; AAG39980.1; -
DR EMBL: AF315681; AAK01367.1; -
DR HSSP: P07987; 1CB2.
DR InterPro: IPR001524; Glyco_hydro_6.
DR Pfam: PF00734; CBD_1; 1.
DR PRINTS: PR00733; GLYDRLASE6.
DR ProDom: PD001821; CBD_fungal; 1.
DR ProDom: PD003733; Glyco_hydro_6; 1.
DR SMART: SM00236; fCBD; 1.
DR PROSITE: PS00562; CBD_FUNGAL; 1.
KM Hydrolase.
SQ SEQUENCE 471 AA; 49712 MW; 45FF20AE836AD3B0 CRC64;

Query Match 24.5%; Score 552; DB 3; Length 471;
Best Local Similarity 33.8%; Pred. No. 2e-31;
Matches 146; Conservative 51; Mismatches 153; Indels 82; Gaps 13;

QY 1 ATHVDNPGATFEVNPMAQEOSEANQTNATLAAKRVSTSTAVMMD----- 52
DB 110 AYSGNPVGVTWPMANAYASVSLALPSLGMATAAATAKVPFWMDTLTKPTLM 169
QY 53 --RIAINGVNGSGPLTLYLDAALSOQOGTTPVEVIEIYIDLPGDCALASNGELPATA 110
DB 170 EQLTADITANKNG--NYAG-----QFVYVDLPDRDCAALASNGEYSIAD 213
QY 111 AGLOTYETQYIDPISLISNPKYSRLVITIEPDSLPAVYTNMSIQCATVAPVYEGGI 170
DB 214 GGVAKYK-NYIDITRQIV--VEYSIRILLVIEPDSLAVTNLTSPKCANQASYLECT 270
QY 171 EYALTKLAIPNVYLYMDAAHSGMLGMPNNAAGYQVQKYLNASIGVNGIDGFYTNAN 230
DB 271 NYAVYQLN-LPVMAYMLDAGHAGMLGWPNADPAQALFANVYKNASPRALNGLATNVAN 329
QY 231 YTPLEKPEMTATQVGGQPVESANFYQNNPDIIDEADYAVDLYSRLVAAGFPSSIGMLIDT 290
DB 330 YN-----GWNITSPSYTOGNNAVYNEKLY-IHALGRILANHGWSNAFFITDQ 375
QY 291 LRNGWGNEPTGPSTADVNTFVNSKIDLRQHRGLMCNONGAGLGPQASPTDFPNA 350
DB 376 GRSG-----KQPLG-----QQQWDMCNVIGTGFGRPSANTD--S 410
QY 351 HLDAYVWIKPESDGTSAASDPPTGKKSDDPCDPTTYSYGLTNALPNSPIAGOWPEA 410
DB 411 LLDSTVWVKPGEGDGTSDSSAP---RFDShc-----ALPDAQAPAAQAMWQA 457
QY 411 QPDQLVANARRA 422
DB 458 YFVOLLTNANPS 469

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLOBIOHYDROLASE II.
GN ACC2.
OS Acremonium cellulolyticus.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Acremonium.
OX NCBI_TaxID=87064;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Y-94;
RA Watanabe M.;
RT "Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases."
RL EMBL: AB022429; CAB19780.1; -
DR HSSP: P07987; 1CB2.
DR InterPro: IPR00254; CBD_fungal.
DR InterPro: IPR001524; Glyco_hydro_6.
DR Pfam: PF00734; CBD_1; 1.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PR00733; GLYDRLASE6.
DR ProDom: PD001821; CBD_fungal; 1.
DR ProDom: PD003733; Glyco_hydro_6; 1.
DR SMART: SM00236; fCBD; 1.
DR PROSITE: PS00562; CBD_FUNGAL; 1.
DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_2; UNKNOWN_1.
KM Hydrolase.
SQ SEQUENCE 457 AA; 47782 MW; 571205A393D760CE CRC64;

Query Match 24.0%; Score 539; DB 3; Length 457;
Best Local Similarity 34.0%; Pred. No. 1.6e-30;
Matches 143; Conservative 56; Mismatches 157; Indels 64; Gaps 14;

QY 6 NPVAGATFEVNPMAQEOSEANQTNATLAAKRVSTSTAVMMDRIAINGVNGPG 65
DB 99 NFPSGQLYANAYYSSEVHTLALPSLGSLLAAATKAKEIPFVLDTRAKY-----PT 152
QY 66 LTTYL-DAALSOQOGTTPVEVIEI-VIYDLPGDCALASNGELPATAAGLOTYETQYIDP 123
DB 153 MGTVLANTIEANKAGASPIINGIFVYVDLPDRDCAALASNGEYTVANNGVANYKA-YIDS 211
QY 124 IASLISNPKYSRLVITIEPDSLPAVYTNMSIQCATVAPVYEGGIEYALTKLAIPNV 183
DB 212 IYAOLK--AYPVHTILLIEPDSLAVTNLTSTAKCAEQASVYECVNTALINLN-LNAV 268
QY 184 YLYMDAAHSGMLGMPNNAAGYQVQKYLNASIGVNGIDGFYTNANTPLKEPMTATQ 243
DB 269 ANYIDAGHAGMLGWSANLSPAQALFATVYKNASAPASJGLATNVANYN-----AW 319
QY 244 QVGGQPVESANFYQNNPDIIDEADYAVDLYSRLVAAGFPSSIGMLIDTLRNGWGNEPTG 303
DB 320 STSSPP---SYTSGDSNDEKLYINALSPLLTSSNGWPA--HFIMDTSRNG---VQPT- 369
QY 304 PSTADVNTFVNSKIDLRQHRGLMCNONGAGLGPQASPTDFPNAHLDAYVWIKPGE 363
DB 370 -----KQAPGDMCNVIGTGFQVQPTNTICD---PLEDAFVWVKPGE 409
QY 364 SDGTSAASDPPTGKKSDDPCDPTTYSYGLTNALPNSPIAGOWPEA 423
DB 410 SDGTSNNS---ATRYDFHCG-----YSDALQAPAEACTWQAFVOLLTNANPAL 456

Search completed: August 29, 2002, 16:37:45
Job time: 955 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:22:32 ; Search time 47.58 Seconds

(without alignments)
217.151 Million cell updates/sec

Title: US-09-917-384-4
Perfect score: 2249
Sequence: 1 AHWVNPYAGATFFVNPYWA.....ACQWPAQFDQLVANRPAY 423

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	639.5	28.4	352	US-09-286-691-26	Sequence 26, Appl
2	639.5	28.4	352	US-09-687-147-26	Sequence 26, Appl
3	555	24.7	360	US-09-286-691-25	Sequence 25, Appl
4	555	24.7	360	US-09-687-147-25	Sequence 25, Appl
5	553.5	24.6	461	US-09-118-319-9	Sequence 9, Appl
6	545	24.0	470	US-09-118-319-8	Sequence 8, Appl
7	539	24.0	457	US-09-142-759-1	Sequence 1, Appl
8	538	23.9	360	US-09-286-691-24	Sequence 24, Appl
9	538	23.9	360	US-09-687-147-24	Sequence 24, Appl
10	538	23.9	365	US-08-169-948B-12	Sequence 12, Appl
11	538	23.9	365	US-08-448-873-12	Sequence 12, Appl
12	538	23.9	365	US-08-382-452D-12	Sequence 12, Appl
13	439.5	19.5	449	US-09-118-319-7	Sequence 7, Appl
14	439.5	19.5	449	US-09-286-691-4	Sequence 4, Appl
15	439.5	19.5	449	US-09-687-147-4	Sequence 4, Appl
16	436	19.4	428	US-09-118-319-5	Sequence 5, Appl
17	434.5	19.3	326	US-09-286-691-23	Sequence 23, Appl
18	434.5	19.3	326	US-09-687-147-23	Sequence 23, Appl
19	431.5	19.2	432	US-09-118-319-2	Sequence 2, Appl
20	411	18.3	459	US-09-118-319-6	Sequence 6, Appl
21	411	18.3	459	US-09-286-691-2	Sequence 2, Appl
22	411	18.3	459	US-09-687-147-2	Sequence 2, Appl
23	242.5	10.8	291	US-09-286-691-28	Sequence 28, Appl
24	242.5	10.8	291	US-09-687-147-28	Sequence 28, Appl
25	236.5	10.5	286	US-09-286-691-27	Sequence 27, Appl
26	236.5	10.5	286	US-09-687-147-27	Sequence 27, Appl
27	231	10.3	490	US-09-109-841-2	Sequence 2, Appl

28	193.5	8.6	290	US-09-286-691-29	Sequence 29, Appl
29	193.5	8.6	290	US-09-687-147-29	Sequence 29, Appl
30	104	4.6	632	US-08-443-104-4	Sequence 4, Appl
31	104	4.6	632	US-08-442-859-4	Sequence 4, Appl
32	104	4.6	632	US-08-398-489-4	Sequence 4, Appl
33	104	4.6	632	PCT-US95-05534-4	Sequence 4, Appl
34	103	4.6	363	US-08-978-559A-2	Sequence 2, Appl
35	99.5	4.4	388	US-08-894-772-2	Sequence 2, Appl
36	99.5	4.4	388	US-09-207-844-2	Sequence 2, Appl
37	98	4.4	1042	US-08-928-361B-11	Sequence 11, Appl
38	98	4.4	1837	US-08-928-361B-5	Sequence 5, Appl
39	96.5	4.3	1751	US-09-136-574A-44	Sequence 44, Appl
40	96	4.3	363	US-09-336-601-1	Sequence 1, Appl
41	94	4.2	626	US-08-596-300A-7	Sequence 7, Appl
42	94	4.2	626	US-08-596-300A-14	Sequence 14, Appl
43	94	4.1	1004	US-09-268-347-30	Sequence 30, Appl
44	92.5	4.1	1043	US-08-928-361B-30	Sequence 30, Appl
45	92.5	4.1	1721	US-08-700-651-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-09-286-691-26
Sequence 26, Application US/09286691
Patent No. 6190189
GENERAL INFORMATION:
APPLICANT: Li, Xin-Liang
APPLICANT: Ljungdahl, Lars G.
TITLE OF INVENTION: Cellulases and Coding Sequences
FILE REFERENCE: 42-96
CURRENT APPLICATION NUMBER: US/09/286,691
CURRENT FILING DATE: 1999-04-05
EARLIER APPLICATION NUMBER: US 60/027, 883
EARLIER FILING DATE: 1996-10-04
EARLIER APPLICATION NUMBER: PCT US97/18008
EARLIER FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 352
TYPE: PRT
ORGANISM: Agaricus bisporus
US-09-286-691-26

Query Match 28.4%; Score 639.5; DB 4; Length 352;
Best Local Similarity 36.6%; Pred. No. 1.1e-56;
Matches 153; Conservative 62; Mismatches 134; Indels 69; Gaps 12:
QY 6 NPYAGATFFVNPYMAOYSEANOTNATLAKMRVSTYSTAVWMDRIALINGVGGPG 65
||| | : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 4 NPYTKTWLSPFYADEVAQAADISNPSLATKASVAKIPFWFDIVAKV-----PD 57
QY 66 LTTVDALDSQGGTTPYIEIVYIDLPGRDCAALASNGELPATAGLOTQYTOYIDPA 125
||| | : : : : : ||| | : : : : : ||| | : : : : : ||| | : : : : :
DB 58 LGQVLADRSKNQ-----LVQIVYIDLPGRDCAALASNGELPATAGLOTQYTOYIDPA 111
QY 126 SLTSPKYSRLIVYIEPDSLPNAVNTNMSIOACATAVPYEGGLEATLTKLHAIPNYI 185
: : : : : ||| | : : : : : ||| | : : : : : ||| | : : : : : ||| | : : : : :
DB 112 AQIK--QPPDVSVAVIEPDSLANVLTNINVOGKCANQASAYKEGYIYAVQKNAV-GVYM 168
: : : : : ||| | : : : : : ||| | : : : : : ||| | : : : : : ||| | : : : : :
QY 186 YMDAHSGLGKPNNAAGVQVQKVLNASTVNGIDGFTVTANTYTPKEPMTATQOV 245
||| | : : : : : ||| | : : : : : ||| | : : : : : ||| | : : : : :
DB 169 YIDACHAGWIGPAMISPAQQLFAQIYRDGSPRLRGATVAVAFNMLRAS----- 220
QY 246 GGQPPESANFYQWNPDIIEADYAVDLSTLYAAGPSSIGMLIDTLRNGWGPNEPTGPS 305
: : : : : ||| | : : : : : ||| | : : : : : ||| | : : : : : ||| | : : : : :
DB 221 SPDPIT-----QGSNYDEIRHIEALAPMLSNAGFPAR--FVVDGRSG----- 262
QY 306 TATDVNTEFVNSKIDLRHRIAGMCNONGAGLQPPQASPFPDPNHLDAVYVWIKPGESD 365

Db 263 -----VONTROGDMCNKAGFGORP---TTNGSSLIDIAVWKKPGGED 307
Qy 366 GTSASDPTTGKSDPMCDPTTYSYGVLTNALPNSPIAGOWFPAGFDOLVANARPAY 423
Db 308 GTSNDSSP---RFDSHCS-----LSDAHQPAPEAGTWFQAFETVLVANANPAL 352

RESULT 2

US-09-687-147-26
; Sequence 26, Application US/09687147
; Patent No. 6268198
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96a
; CURRENT APPLICATION NUMBER: US/09/687,147
; CURRENT FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/027,883
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: PCT US97/18008
; PRIOR FILING DATE: 1997-10-03
; PRIOR APPLICATION NUMBER: 09/286,691
; PRIOR FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 26
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Agaricus bisporus
US-09-687-147-26

Query Match 28.4%; Score 639.5; DB 4; Length 352;
Best Local Similarity 36.6%; Pred. No. 1,1e-56;
Matches 153; Conservative 62; Mismatches 134; Indels 69; Gaps 12;

Qy 6 NYAGATFFVNPYMAQEOSEANOTNATLAKMRVSYSTAVWMDRIAINGVNGFG 65
Db 4 NYTGATVWLSPEYDEVAAADISNPISLATKASVAKIPFEVMDTVAKV-----PD 57
Qy 66 LTTVLDAAISOOGTTPETIEIYVDLPGRDCAALASNGELPATAAGLOTYETQYIDPTA 125
Db 58 LGTYLDAAASKNO-----LVQIVYVDLPDRCALASNGEFLANDGLNKKY-NYDQIA 111
Qy 126 SILSNPKYSLRIYVITIEPDSLPAVNTNMSIOACATAVPYEAGEIEYALTKLHAIPNYI 185
Db 112 AQIK--QFPDVSVAIVIEPDSLANTVNLNVOKCANQSAKYEGVYAVOKLNAV-GYIM 168
Qy 186 YMDAASHGWLGPNNASGYVOEVOKYLNASIGVNGIDGVTNTANTPLKEPMTATQOV 245
Db 169 YIDAGAGLWGPNNASPPAQLFAQIYRDAGSPRNLRGIAITVANANPALRAS----- 220
Qy 246 GQPYESANFYQWNPDIIDEADYAVDLSRLVAAGPPSSIGMLIDTLRNGMGPNPTGPS 305
Db 221 SEDPTI-----QGNSTYDIHITIEALPMLSNAGFPAH--FYDQGRSG----- 262
Qy 306 TATDVNTFVNOGRKIDLRQHRGLMONGAGLQOPQASPTDEPNNAHLDAVWIKPGEED 365
Db 263 -----VONTROGDMCNKAGFGORP---TTNGSSLIDIAVWKKPGGED 307
Qy 366 GTSASDPTTGKSDPMCDPTTYSYGVLTNALPNSPIAGOWFPAGFDOLVANARPAY 423
Db 308 GTSNDSSP---RFDSHCS-----LSDAHQPAPEAGTWFQAFETVLVANANPAL 352

RESULT 3

US-09-286-691-25
; Sequence 25, Application US/09286691
; Patent No. 6190189
; GENERAL INFORMATION:

; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96
; CURRENT APPLICATION NUMBER: US/09/286,691
; CURRENT FILING DATE: 1999-04-05
; EARLIER APPLICATION NUMBER: US 60/027,883
; EARLIER FILING DATE: 1996-10-04
; EARLIER APPLICATION NUMBER: PCT US97/18008
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 25
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Fusarium oxysporum
US-09-286-691-25

Query Match 24.7%; Score 555; DB 4; Length 360;
Best Local Similarity 33.7%; Pred. No. 4.5e-48;
Matches 142; Conservative 64; Mismatches 147; Indels 68; Gaps 14;

Qy 5 DNPYAGATFFVNPYMAQEOSEANOTNATLAKMRVSYSTAVWMDRIAINGV 61
Db 3 DNPYAGVDMANNYRSEVNNLAVPRLSGAKATAAKVADVBSFO--WMDTYDHISLME 59
Qy 62 GGGGLTLYDAAISOOGTTPETIEIYVDLPGRDCAALASNGELPATAAGLOTYETQYI 121
Db 60 -----DTLADIRKANAG--GYAGQFVYVDLPDRCALASNGEFLANDGLNKKY-KY 112
Qy 122 DPTASILSNPKYSLRIYVITIEPDSLPAVNTNMSIOACATAVPYEAGEIEYALTKLHAIP 181
Db 113 AKIKGILQ--YSDTKVILVIEPDSLANTVNLNVOKCAKASAKYELFVYAIKELN-LP 169
Qy 182 NYIYMDAASHGWLGPNNASGYVOEVOKYLNASIGVNGIDGVTNTANTPLKEPMTA 241
Db 170 NWSMYLDAGHGWLGPNNAGPNAIGPAKLYAQLYKDAGRSPRNLRGIAITVANANPAL 220
Qy 242 TQVGGQPYESANFYQWNPDIIDEADYAVDLSRLVAAGPPSSIGMLIDTLRNGMGPNP 301
Db 221 ---GWLKSTKPDYTESNNYDEQR--INAFAPLQAQEGMSNKFIVDQGRSG---KOP 271
Qy 302 TGPSTATDVNTFVNOGRKIDLRQHRGLMONGAGLQOPQASPTDEPNNAHLDAVWIKP 361
Db 272 TG-----QKAGQDMCNKAGTGFGLRPSFTYGD--ALADAFVWVKFG 310
Qy 362 GSEDGTSASDPTTGKSDPMCDPTTYSYGVLTNALPNSPIAGOWFPAGFDOLVANARP 421
Db 311 GSEDGTS-----DTSARKYHCG-----LDDALKPAPEAGTWFQAFETVLVANANPAL 357
Qy 422 A 422
Db 358 S 358

RESULT 4

US-09-687-147-25
; Sequence 25, Application US/09687147
; Patent No. 6268198
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96a
; CURRENT APPLICATION NUMBER: US/09/687,147
; CURRENT FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/027,883
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: PCT US97/18008
; PRIOR FILING DATE: 1997-10-03

PRIOR APPLICATION NUMBER: 09/286,691
PRIOR FILING DATE: 1999-04-05
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 25
LENGTH: 360
TYPE: PRT
ORGANISM: Fusarium oxysporum
US-09-687-147-25

Query Match 24.7%; Score 555; DB 4; Length 360;
Best Local Similarity 33.7%; Pred. No. 4,5e-48;

Matches 142; Conservative 64; Mismatches 147; Indels 68; Gaps 14;

QY 5 DNPYAGATFVNDPYMAOEVOSEAA---NOTNATLAAKMRVSTYSTAAMDRITAAINGVN 61
DB 3 DNPYAGVDLMANNYRSEVMNLAVPKLSGAKATAAAKAVADVSFO---WMDTYDHSLSME 59
QY 62 GGPGLTTYLDAAALSOOGGTPPEVIEIYVLDLPGDCALASNGELPATAAGLOTTEYOYI 121
DB 60 -----DTLADIRKANAG-GKYAGQFVYVDLPRDCAAAASNGEYSLDGDANKYKA-YI 112
QY 122 DPASLSPKYSRLRVTIIEPDSLPAVATNMSIQACATAPVYEOGIEYALTKLHAIP 181
DB 113 AKIKGLQN--YSDTFVILVIEPDSLANTLNTLVNDKCAKESAYKELTYAIKELN-LP 169
QY 182 NYIYMDAHSGLGMPNNASGYOEVOKVLNASIGVNGIDGFTVTANTYTPLEKPFMTA 241
DB 170 NYSMTLDAGHGLGMPANIGPAKLYAQIYKDAKPSRVRGLVTIVSNVN----- 220
QY 242 TQVGGQPVESANFYQWNPIDADYAVDLYSLVAAGFPSSIGMLDITLRNGGPNP 301
DB 221 ---GKMLSTRKPDYTESNPYDEORY-INAFAPLQOEKSNKFTVDGRSG---KQP 271
QY 302 TGPSTATDVNTFVNOGSKIDLROHGLMCNONGAGLGPQASPTDPNNAHLDAVYWKIP 361
DB 272 TG-----OKAOGDMCAKGTGFLRSTNTGD--ALADAFVWVKPG 310
QY 362 GESDGTSAASDPTTGKSDPMDPTTTSYGVLTNALPNSPIAGOWFPAQFDLVANARP 421
DB 311 GESDGT---DTSAAKYVHCG-----LDLAKRPAEAGTWQAFQKOLLNANP 357
QY 422 A 422
DB 358 S 358

RESULT 5

US-09-118-319-9
Sequence 9, Application US/09118319
Patent No. 6114158

GENERAL INFORMATION:

APPLICANT: LI, Xin-Liang

APPLICANT: Chen, Huizhong

APPLICANT: Ljungdahl, Lars G.

FILE OF INVENTION: Orpinomyces Cellulase Celf Protein and Coding Sequences

CURRENT APPLICATION NUMBER: US/09/118,319

CURRENT FILING DATE: 1998-07-17

NUMBER OF SEQ ID NOS: 9

SOFTWARE: Patent Ver. 2.0

SEQ ID NO 9

LENGTH: 461

TYPE: PRT

ORGANISM: Fusarium oxysporum

Query Match 24.6%; Score 553.5; DB 3; Length 461;
Best Local Similarity 33.5%; Pred. No. 9,6e-48;
Matches 141; Conservative 65; Mismatches 146; Indels 69; Gaps 14;

QY 5 DNPYAGATFVNDPYMAOEVOSEAA---NOTNATLAAKMRVSTYSTAAMDRITAAINGVN 61
DB 105 DNPYAGVDLMANNYRSEVMNLAVPKLSGAKATAAAKAVADVSFO---WMDTYDHSLSME 161
QY 62 GGPGLTTYLDAAALSOOGGTPPEVIEIYVLDLPGDCALASNGELPATAAGLOTTEYOYI 121
DB 162 -----DTLADIRKANAG-GKYAGQFVYVDLPRDCAAAASNGEYSLDGDANKYKA-YI 214
QY 122 DPASLSPKYSRLRVTIIEPDSLPAVATNMSIQACATAPVYEOGIEYALTKLHAIP 181
DB 215 AKIKGLQN--YSDTFVILVIEPDSLANTLNTLVNDKCAKESAYKELTYAIKELN-LP 271
QY 182 NYIYMDAHSGLGMPNNASGYOEVOKVLNASIGVNGIDGFTVTANTYTPLEKPFMTA 241
DB 272 NYSMTLDAGHGLGMPANIGPAKLYAQIYKDAKPSRVRGLVTIVSNVN----- 322
QY 242 TQVGGQPVESANFYQWNPIDADYAVDLYSLVAAGFPSSIGMLDITLRNGGPNP 301
DB 323 ---GKMLSTRKPDYTESNPYDEORY-INAFAPLQOEKSNKFTVDGRSG---KQP 373
QY 302 TGPSTATDVNTFVNOGSKIDLROHGLMCNONGAGLGPQASPTDPNNAHLDAVYWKIP 361
DB 374 TG-----OKAOGDMCAKGTGFLRSTNTGD--ALADAFVWVKPG 412
QY 362 GESDGTSAASDPTTGKSDPMDPTTTSYGVLTNALPNSPIAGOWFPAQFDLVANARP 421
DB 413 GESDGT---DTSAAKYVHCG-----LDLAKRPAEAGTWQAFQKOLLNANP 458
QY 422 A 422
DB 459 S 459

RESULT 6

US-09-118-319-8
Sequence 8, Application US/09118319
Patent No. 6114158

GENERAL INFORMATION:

APPLICANT: LI, Xin-Liang

APPLICANT: Chen, Huizhong

APPLICANT: Ljungdahl, Lars G.

FILE OF INVENTION: Orpinomyces Cellulase Celf Protein and Coding Sequences

CURRENT APPLICATION NUMBER: US/09/118,319

CURRENT FILING DATE: 1998-07-17

NUMBER OF SEQ ID NOS: 9

SOFTWARE: Patent Ver. 2.0

SEQ ID NO 8

LENGTH: 470

TYPE: PRT

ORGANISM: Trichoderma reesei

US-09-118-319-8

Query Match 24.3%; Score 546; DB 3; Length 470;
Best Local Similarity 33.6%; Pred. No. 5,8e-47;
Matches 145; Conservative 51; Mismatches 154; Indels 82; Gaps 13;

QY 1 ATRVNDPYAGATFVNDPYMAOEVOSEANOTNATLAAKMRVSTYSTAAMDRITAAINGVN 52
DB 109 ATRVNDPYAGATFVNDPYMAOEVOSEANOTNATLAAKMRVSTYSTAAMDRITAAINGVN 168
QY 53 -RIAAINGVNGPGLTTYLDAAALSOOGGTPPEVIEIYVLDLPGDCALASNGELPATA 110
DB 169 EOTLADIRKANAG--NTAG-----QPVYVDLPRDCAAAASNGEYSLAD 212
QY 111 AGLOTTEYOYIDPISLSPKYSRLRVTIIEPDSLPAVATNMSIQACATAPVYEOGIEYALTKLHAIP 170
DB 213 GGYAKYK-NTIDIRQIV--VEYSDIRTLVIEPDSLANTLNTLVNDKCAKESAYKELTYAIKELN-LP 269
QY 171 EYALTKLHAIPNYIYMDAHSGLGMPNNASGYOEVOKVLNASIGVNGIDGFTVTANTYTPLEKPFMTA 230
DB 270 NYAVTOLN-LPNNAYMIDAGHGLGMPANODPAQDLFANVYKNASSPRALRGLATNIVAN 328

QY	231	YTPLEKPEMTATQOYGGPVESANFYQNNPDIADYAVDLYSRVAAGFSSIGMLTD	290
Db	329	YN-----GNMTSPSPSTTQGNAYVNEKLYTHALGPLLANNGW-SNAFFITDQ	374
QY	291	LRNGMGNEPTGPSTADVTNVNOSKIDLRQHRGLMCNONGAGLGOPPOASPTDFPNA	350
Db	375	GRSG--KQPLTG-----QQQMGDCNVIIGTGFGRPSANTGD---S	409
QY	351	HLDAVWVTKKPGESDGTSAASDPTTGKKSDDPCDPTTYSIGVLTNALPNSIACQWPPA	410
Db	410	LIDSFVWVKPGEGECGTSDSSAP-----RFDSHC-----ALPDALQAPQAGAMFOA	456
QY	411	QPDQVLANAPRA 422	
Db	457	YFVQVLLTNANPS 468	
RESULT 7			
US-09-142-759-1			
Sequence 1, Application US/09142759A			
Patent No. 6127160			
GENERAL INFORMATION:			
APPLICANT: YAMANABE, Takashi			
APPLICANT: WATANABE, Manabu			
APPLICANT: HAMAYA, Toru			
APPLICANT: SUMIDA, Naomi			
APPLICANT: AOYAGI, Kaoru			
APPLICANT: MURAKAMI, Takeshi			
TITLE OF INVENTION: PROTEIN HAVING CELLULOSE ACTIVITY AND PROCESS FOR			
FILE OF INVENTION: PRODUCING THE SAME			
FILE REFERENCE: Q51673			
CURRENT APPLICATION NUMBER: US/09/142,759A			
CURRENT FILING DATE: 1998-09-14			
EARLIER APPLICATION NUMBER: PCT/J997/00824			
EARLIER FILING DATE: 1997-03-14			
EARLIER APPLICATION NUMBER: P. HEI-8-84479			
EARLIER FILING DATE: 1996-03-14			
NUMBER OF SEQ ID NOS: 30			
SOFTWARE: Patent In Ver. 2.1			
SEQ ID NO 1			
LENGTH: 457			
TYPE: PRT			
ORGANISM: Acetomoniun cellulolyticus			
US-09-142-759-1			
Query Match			
Best Local Similarity 34.0% Score 539; DB 3; Length 457;			
Matches 143; Conservative 56; Mismatches 157; Indels 64; Gaps 14;			
QY	6	NPYAGATFFNVYVMAQEOSEANQTNATLAKMNVVSTSYSAVMMDRITAAINGVNGPG	65
Db	99	NFSGSYQLYANAYISEVHTLAIPLSLGSLAAATKKAETPFVWLDRPAK-----PT	152
QY	66	LTYYV-DAAISOOGQGTTPVEVIEI-VIYDIPGRDCAALASNGELPATPAAGLQTYETQYIDP	123
Db	153	MGTVALNTEAANKAKASPGIAGIFVAVYDIPDRDCAALASNGEYTVANNQVANYKA-YIDS	211
QY	124	IASISNPKYSSLRIVTITEPDSLSENAVYNNWSIQACATAVPYEEOGIEVALKRLAIPRY	183
Db	212	IYAOLK--AYPDVHTLITIEPDSLANNMTNLSLSTACAEAOASAYFECVNTALINLN-LANV	268
QY	184	YLYMDAAHSGWLGMPPNNAAGVQEOQVULNASIGVNGIDGFVYNTANTYPLKEPMTATQ	243
Db	269	AMYIDAGHAGWLGWGSANLSPAQDLATYATYKKNASAPASLRLGLATVNAVNTY-----AM	319
QY	244	OYGGOPVESANFYQNNPDIADYAVDLYSRVAAGFSSIGMLIDTFLRNGMGNEPTG	303
Db	320	SISSP-----SYTSDSNVDEKLYINALSLPLTTSNGWPA--HFIMDTSRNG-----VQPT-	369
QY	304	PSTANDVMTFVNVQSKIDLRQHRGLMCNONGAGLGOPPOASPTDFPNAHLDAVWVTKKPE	363

```

Db 370 -----KQAMGDMCNVIGTGFVGQPTNTGDD--PLEDAFVWVKPGGE 409
OY 364 SDGTSASADPTTKGKSDPMCDPTTTSIGVLTNALPNSPIAGQFPAPQFDLVANARAV 423
Db 410 SDGTSNSS-----ATRYDFHCG-----YSDALQAPAEAGTWOAYFVOLLTNANPAL 456

RESULT 8
US-09-286-691-24
; Sequence 24, Application US/09286691
; Patent No. 6190189
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Lj, Xunzhong
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96
; CURRENT APPLICATION NUMBER: US/09/286,691
; CURRENT FILING DATE: 1999-04-05
; EARLIER APPLICATION NUMBER: US 60/027,883
; EARLIER FILING DATE: 1996-10-04
; EARLIER APPLICATION NUMBER: PCT US97/18008
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-286-691-24

Query Match 23.9%; Score 538; DB 4; Length 360;
Best Local Similarity 33.5%; Pred. No. 2.4e-46;
Matches 143; Conservative 50; Mismatches 152; Indels 82; Gaps 13;

OY 6 NPVYGAFFGVNRYMQEVSSEANQNTNNTLAKMRYSTISAAVMD-----RIA 55
Db 4 NPEVGVTVWANNAYVASEVSSLAIPSLTGAMATAAAAAAAKVPSEFMDLDTLTKPTIMEQTILA 63
OY 56 AINGVNGSGPGLTYYIDAALSOQOQTTPVEVIEIVYDLPGRCOCALASNGELPATAAGLOT 115
Db 64 DIRTKNKG--NYAG-----QPVYDLEDRDCALASNGEYSIADGVAK 107
OY 116 YETQYIDDIASILSNPKYSLSRIYTIIEBDSLPPNAVNTNMSIOACATAVPYEEOGIEVALT 175
Db 108 YK-NYIDTRIQLIV--VEYSIDRTILYIEBDSLANTVTLNGLPFCANAGSAVLECIYAVT 164
OY 176 KLHAIPNNYITMDAAHSGMLGMPNNAAGSYQVEQAVYLNASIGVNGIDGFYNTANTYTPLK 235
Db 165 QLN-LPNNVAMYLDAHGMWLPAPADPAQOLFANVYNNASSPRALRGATVNTVANN-- 220
OY 236 EPPFMATQOVGQPVESANFEYQWNPNDIDEDADYADVLYLSRLVAAGFPSSIGMLIDTLRNGW 295
Db 221 -----GNITSPESYTOGNANVYIEKLYIHAIGPLANHGM-SNAFFTTDGRSG- 268
OY 236 GGPNEPTGSPSTATDVNTEVNOSKIDLRQHRGLMCNONGAGLGQAPQASPTDFPNALDAY 355
Db 269 ---KQPTG-----QQQWDMCVIGTGFGRSANTMGD---SLDSF 304
OY 356 VWIKRPGESDGTSAASDPPTTKGKSPMCDPTTTSIGVLTNALPNSPIAGQFPAPQFDOL 415
Db 305 VWVKGGBGECDDSDSAP---RFDSHC-----ALPDALQAPQAPQAGAFQAYFVOL 351
OY 416 VANARPA 422
Db 352 LFNANPS 358

RESULT 9
US-09-687-147-24
; Sequence 24, Application US/09687147
; Patent No. 6268198

```

```

; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96a
; CURRENT APPLICATION NUMBER: US/09/687,147
; CURRENT FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/027,883
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: PCT US97/18008
; PRIOR FILING DATE: 1997-10-03
; PRIOR APPLICATION NUMBER: 09/286,691
; PRIOR FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentl Ver. 2.0
; SEQ ID NO 24
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Trichoderma reesei
; US-09-687-147-24

```

```

Query Match      23.9%; Score 538; DB 4; Length 360;
Best Local Similarity 33.5%; Pred. No. 2.4e-46;
Matches 143; Conservative 50; Mismatches 152; Indels 82; Gaps 13;

```

```

OY 6 NPVAGATFFVNPYMAOEVOSEANOTNATLAKMRYVSTYSTAVMD-----RTA 55
DB 4 NPVGVGTPPANKYAYSEVSLAIPSLTGAMATPAAVAAPVPSMWLDTDKTPLMGTTLA 63
OY 56 AINGVNGPGLTTYLDALSOOGCTTPEVIEIYIDLPGRCALASNGELPATAAGLOT 115
DB 64 DIRTARKNG--NYAG-----QFVYDLPDRDCAALASNGEYSIADGVAK 107
OY 116 YEQYIDPILASISNPKYSLRYITIEPDSLPAVNTNMSIQACATAVPYEGIEYALT 175
DB 108 YK-NYIDTRIQLV--VEYSDIRTLVIEPDSLANTLTNLGTPCANAQSAIYECINYAVT 164
OY 176 KLHAIPNYIYMDAHSGLGMPNNAAGYVOEVOKVLNASIGVNGIDGFVTNTANTPLK 235
DB 165 QLN-LPNVAMYLDAGHAGMLGWPANODPPAQLFANVYKNASPRALRGLATNVANIN--- 220
OY 236 EPPMATQOVGQPVESANFYQWNPDIADYADVLSRLVAAGPSSIGMLIDTLRNGW 295
DB 221 -----GNWITSPSYTOGNNAVYNEKLYIHAIIGPLANHG--SNAFFITDGRSG- 268
OY 296 GGPNEPTGSTATDVNTFVNOGRKIDLRHRLGMCNONGAGLGOPPOASPTDFPNALDAY 355
DB 269 ---KQPTG-----QOQWGMCMNVIGTGFIRPSANTGD---SLDSE 304
OY 356 VWIKPGEISDGTSAASDPTTGKSDPMCDPTVYTSYGLTNALPNSPIAGOWPFAQFDOL 415
DB 305 VWIKPGEISDGTSAASDPTTGKSDPMCDPTVYTSYGLTNALPNSPIAGOWPFAQFDOL 415
OY 416 VANARPA 422
DB 352 LTNANPS 358

```

```

RESULT 10
US-08-169-948B-12
; Sequence 12, Application US/08169948B
; Patent No. 5861271
; GENERAL INFORMATION:
; APPLICANT: Fowler, Timothy
; APPLICANT: Ward, Michael
; APPLICANT: Clarkson, Kathleen
; APPLICANT: Collier, Katherine
; APPLICANT: Larens, Edmund
; TITLE OF INVENTION: No. 5861271el Cellulase Enzymes and Systems
; TITLE OF INVENTION: For Their Expression
; NUMBER OF SEQUENCES: 48

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentl Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,948B
; FILING DATE: DEC 17 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A.
; REGISTRATION NUMBER: 33,401
; REFERENCE/DOCKET NUMBER: GC226
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7536
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 365 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-169-948B-12

```

```

Query Match      23.9%; Score 538; DB 2; Length 365;
Best Local Similarity 33.3%; Pred. No. 2.5e-46;
Matches 144; Conservative 51; Mismatches 155; Indels 82; Gaps 13;

```

```

OY 1 ATHVNPVAGATFFVNPYMAOEVOSEANOTNATLAKMRYVSTYSTAVMD----- 52
DB 4 ATYSGNPFVGVTPPANKYAYSEVSLAIPSLTGAMATPAAVAAPVPSMWLDTDKTPLM 63
OY 53 --RIAINGVNGPGLTTYLDALSOOGCTTPEVIEIYIDLPGRCALASNGELPATA 110
DB 64 EOTLADIRANKNG--NYAG-----QFVYDLPDRDCAALASNGEYSIAD 107
OY 111 AGLOTETQYIDPILASISNPKYSLRYITIEPDSLPAVNTNMSIQACATAVPYEGOI 170
DB 108 GYVAKYK-NYIDTRIQLV--VEYSDIRTLVIEPDSLANTLTNLGTPCANAQSAIYECI 164
OY 171 EYALTRKLHAIPNYIYMDAHSGLGMPNNAAGYVOEVOKVLNASIGVNGIDGFVTNTAN 230
DB 165 NYAVTQLN-LPNVAMYLDAGHAGMLGWPANODPPAQLFANVYKNASPRALRGLATNVAN 223
OY 231 YTLKEPFMATQOVGQPVESANFYQWNPDIADYADVLSRLVAAGPSSIGMLIDT 290
DB 224 YN-----GNWITSPSYTOGNNAVYNEKLYIHAIIGPLANHG--SNAFFITDQ 269
OY 291 LRNGMGGPNEPTGSTATDVNTFVNOGRKIDLRHRLGMCNONGAGLGOPPOASPTDFPN 350
DB 270 GRSG---KQPTG-----QOQWGMCMNVIGTGFIRPSANTGD---S 304
OY 351 HLDAYVWIKPGEISDGTSAASDPTTGKSDPMCDPTVYTSYGLTNALPNSPIAGOWPFA 410
DB 305 LLDSEFVWIKPGEISDGTSAASDPTTGKSDPMCDPTVYTSYGLTNALPNSPIAGOWPFA 410
OY 411 QFDOLVANARPA 422
DB 352 YFVOLLTNANPS 363

```

```

RESULT 11
US-08-448-873-12
; Sequence 12, Application US/08448873
; Patent No. 5874276

```

```

1  GENERAL INFORMATION:
2  APPLICANT: Fowler, Timothy
3  APPLICANT: Ward, Michael
4  APPLICANT: Clarkson, Kathleen
5  APPLICANT: Collier, Katherine A.
6  APPLICANT: Larenas, Edmund
7  TITLE OF INVENTION: No. 5874276el Cellulase Enzymes and Systems
8  TITLE OF INVENTION: For Their Expressions
9  NUMBER OF SEQUENCES: 48
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: Genencor International
12 STREET: 180 Kimball Way
13 CITY: South San Francisco
14 STATE: CA
15 COUNTRY: USA
16 ZIP: 94080
17
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: Patentln Release #1.0, Version #1.25
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/448, 873
25 FILING DATE:
26 CLASSIFICATION: 435
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: 08/169, 948
29 FILING DATE: 17-DEC-1993
30 ATTORNEY/AGENT INFORMATION:
31 NAME: Stone, Christopher L.
32 REGISTRATION NUMBER: 35,696
33 REFERENCE/DOCKET NUMBER: GC226D14
34 TELECOMMUNICATION INFORMATION:
35 TELEPHONE: (415) 742-7555
36 TELEFAX: (415) 742-7217
37 INFORMATION FOR SEQ ID NO: 12:
38 SEQUENCE CHARACTERISTICS:
39 LENGTH: 365 amino acids
40 TYPE: amino acid
41 TOPOLOGY: linear
42 MOLECULE TYPE: protein
43
44 US-08-448-873-12
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
9
```

[illegible]

Db 224 YN-----GWNITSPSTOGNAYNEKLYIHAIIGCLLNNHGN-SNAFFITDQ 269
Qy 291 LRNGMGPEPTGPSTATDVNTFVNOSKIDLROHRLMCNONGAGLGPPOASPTDFPNA 350
Db 270 GRSG-----KQPTG-----QQQMGDMCNVIGTGIRGINSANTGD---S 304
Qy 351 HLDAYVWTPRPGESDGTSAASDPTTGKSDPMCDPTTYSYGLVNALPNSPIAGWFPFA 410
Db 305 LLDSEFWVRPGECDDGTSDSSAP---RDSHC-----ALPDALQAPQAGAMFQA 351
Qy 411 OFDOLVANARPA 422
Db 352 YFVQLLTNANPS 363

RESULT 13
US-09-118-319-7
Sequence 7, Application US/09118319
Patent No. 6114158

GENERAL INFORMATION:
APPLICANT: Li, Xin-Liang
APPLICANT: Chen, Huizhong
APPLICANT: Ljungdahl, Lars G.
TITLE OF INVENTION: Orlinomyces Cellulase Celf Protein and Coding Sequences
FILE REFERENCE: 33-98sequence 1stling
CURRENT APPLICATION NUMBER: US/09/118,319
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patentlin Ver. 2.0
SEQ ID NO 7
LENGTH: 449
TYPE: PRT
ORGANISM: Orlinomyces sp. PC-2
US-09-118-319-7

Query Match 19.5%; Score 439.5; DB 3; Length 449;
Best Local Similarity 32.0%; Pred. No. 3.8e-36;
Matches 136; Conservative 55; Mismatches 131; Indels 103; Gaps 18;

Qy 2 THVDNRYAGATFFVNP-----YMAOEYQSEANOTNATLAKMRVSTYSTAVMMDRIMAI 57
Db 121 TVVNSLPTSDNPFENELYSNFKFQGEVDOSIORLSGLQEKAKKVVYVPAAMLANSAGAT 180
Qy 58 NGVNGSGPLTYYDAALSOQOQTTPEVIEIYIDLGRDCAALASNGELPATAGLOTYE 117
Db 181 NEV-----ARYLNEAGS-----KTVVFVLYMPTTRDCNAGSGNG---GADNLSTYQ 223
Qy 118 TOYIDPIASILSNPKYSSLRIYTIIEPDSL.PNAVITNMSIOACATAVPEYEGIEVALTKL 177
Db 224 -GVNSIYMTIN--QYPSRIYMIIEPDTIGNLVTANNAN--CNAVDMHKQALSTYAIKRF 279
Qy 178 HAIPNYIYMDAHSGLWGLPNNASGYVOEQVQLNASIGVNCIDGFVTNTANTYPLKEP 237
Db 280 GTQKNRVYLDAAHGWL--NSSADRTAEVIAEILRNAGNGKIRGISITVSNY----- 330
Qy 238 FMTATQOVGQPYESANFYQWNPDIIDEADYAVDLYSRLVAAGFPSSIGMLIDTLRNGWG 297
Db 331 -----QPYISE--YQYHONLNR-----LESRGV-----RGMKFIYDTSRNG--- 365
Qy 298 PNEPTGPSTATDVNTFVNOSKIDLROHRLMCNONGAGLGPPOASPTDFPNAH-LDAYV 356
Db 366 -----RNPSSAT-----WCLKAGAGLQARQANDP--PNNPPLDAYV 400
Qy 357 WIKRPEESDGTSAASDPTTGKSDPMCDPTTYSYGLVNALPNSPIAGWFPFAQFDOLV 416
Db 401 WIKRPEESDASASA-----DPVCRN-----SDSLQAGAPAAAGSMFHDYFVMLL 442
Qy 417 ANARP 421
Db 443 ENANP 447

RESULT 14
US-09-286-691-4
Sequence 4, Application US/09286691
Patent No. 6190189
GENERAL INFORMATION:
APPLICANT: Li, Xin-Liang
APPLICANT: Chen, Huizhong
APPLICANT: Ljungdahl, Lars G.
TITLE OF INVENTION: Cellulases and Coding Sequences
FILE REFERENCE: 42-96
CURRENT APPLICATION NUMBER: US/09/286,691
CURRENT FILING DATE: 1999-04-05
EARLIER APPLICATION NUMBER: US 60/027,883
EARLIER FILING DATE: 1996-10-04
EARLIER APPLICATION NUMBER: PCT US97/18008
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentlin Ver. 2.0
SEQ ID NO 4
LENGTH: 449
TYPE: PRT
ORGANISM: Orlinomyces sp. PC-2
US-09-286-691-4

Query Match 19.5%; Score 439.5; DB 4; Length 449;
Best Local Similarity 32.0%; Pred. No. 3.8e-36;
Matches 136; Conservative 55; Mismatches 131; Indels 103; Gaps 18;

Qy 2 THVDNRYAGATFFVNP-----YMAOEYQSEANOTNATLAKMRVSTYSTAVMMDRIMAI 57
Db 121 TVVNSLPTSDNPFENELYSNFKFQGEVDOSIORLSGLQEKAKKVVYVPAAMLANSAGAT 180
Qy 58 NGVNGSGPLTYYDAALSOQOQTTPEVIEIYIDLGRDCAALASNGELPATAGLOTYE 117
Db 181 NEV-----ARYLNEAGS-----KTVVFVLYMPTTRDCNAGSGNG---GADNLSTYQ 223
Qy 118 TOYIDPIASILSNPKYSSLRIYTIIEPDSL.PNAVITNMSIOACATAVPEYEGIEVALTKL 177
Db 224 -GVNSIYMTIN--QYPSRIYMIIEPDTIGNLVTANNAN--CNAVDMHKQALSTYAIKRF 279
Qy 178 HAIPNYIYMDAHSGLWGLPNNASGYVOEQVQLNASIGVNCIDGFVTNTANTYPLKEP 237
Db 280 GTQKNRVYLDAAHGWL--NSSADRTAEVIAEILRNAGNGKIRGISITVSNY----- 330
Qy 238 FMTATQOVGQPYESANFYQWNPDIIDEADYAVDLYSRLVAAGFPSSIGMLIDTLRNGWG 297
Db 331 -----QPYISE--YQYHONLNR-----LESRGV-----RGMKFIYDTSRNG--- 365
Qy 298 PNEPTGPSTATDVNTFVNOSKIDLROHRLMCNONGAGLGPPOASPTDFPNAH-LDAYV 356
Db 366 -----RNPSSAT-----WCLKAGAGLQARQANDP--PNNPPLDAYV 400
Qy 357 WIKRPEESDGTSAASDPTTGKSDPMCDPTTYSYGLVNALPNSPIAGWFPFAQFDOLV 416
Db 401 WIKRPEESDASASA-----DPVCRN-----SDSLQAGAPAAAGSMFHDYFVMLL 442
Qy 417 ANARP 421
Db 443 ENANP 447

RESULT 15
US-09-687-147-4
Sequence 4, Application US/09687147
Patent No. 6268198
GENERAL INFORMATION:
APPLICANT: Li, Xin-Liang
APPLICANT: Ljungdahl, Lars G.
APPLICANT: Chen, Huizhong
TITLE OF INVENTION: Cellulases and Coding Sequences
FILE REFERENCE: 42-96a

```
; CURRENT APPLICATION NUMBER: US/09/687,147
; CURRENT FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/027,883
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: PCT US97/18008
; PRIOR FILING DATE: 1997-10-03
; PRIOR APPLICATION NUMBER: 09/286,691
; PRIOR FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Orpinomyces sp. PC-2
; US-09-687-147-4
```

```
Query Match 19.5%; Score 439.5; DB 4; Length 449;
Best Local Similarity 32.0%; Pred. No. 3.8e-36;
Matches 136; Conservative 55; Mismatches 131; Indels 103; Gaps 18;
```

```
QY 2 THVDNPTAGATFEVNP---YMAQEVQSEANQTNATLAKMRYVSTYSTAVMMDRINAI 57
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 121 TTVNSLPTSDNFENELYSNKKFQGEVDQSIQRLSGLSQEKAKKYYPTAAILAMSGAT 180
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 58 NGVNGGPGCLTTYLDAAALSQOQGTPEVIEIYIDLPGRDCAALASNGELPATAAGLOTYE 117
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 181 NEV-----ARYLNEAGS-----KTVFVFLYMIPTRDCAAGSNG---GADNLSITYQ 223
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 118 TOYIDPIASILSNPKYSLRIVTITIEPDSLPAATNMSIQACATAVPYEGGIEYALFKL 177
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 224 -GYNSTINTIN--QYNSRIYMIIEPDTIGNLTANNAN--CRNVHDMHKQALSTAIKSF 279
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 178 HAIFNVYIYMDAHSGLWGPNNASGVYQVQKYLMASIGVNGIDGFVTNTANTYPLKEP 237
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 280 GTOKNVRVYLDAAHGWL---NSSADRTAEVIAEILRNAGNGKIRGISTVSNY----- 330
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 238 FMTATQOVGGPVSANFYQNNPDIDEADYAVDLXSRLYAAGFPSSIGMLIDTLRNGWG 297
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 331 -----OPVYSE--YQYHQLNRA-----LESRGV-----RGMKEIVDTSRNG--- 365
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 298 PNEPTGPTATDVNTFVNSKIDLRHRLWCNONGAGIGOPPOASPTDPFNAH-IDAYV 356
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 366 -----RNPSSAT-----WCNLRGAGLGARPOANPD--PNMPLLDAYV 400
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 357 WIKPPGESDGTSAASDPTTGKSDPMCDPTYTTSYGVLTJNALPNSPIAGOWFPAQFDQLV 416
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 401 WIKTPGESDSASA-----DPCVCRN-----SDSLQGAPAGSWFHDYFVWLL 442
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 417 ANARP 421
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 443 ENANP 447
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

Search completed: August 29, 2002, 16:22:34
Job time: 364 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:21:42 ; Search time 127 Seconds

(without alignments)
131.190 Million cell updates/sec

Title: US-09-917-384-5

Perfect score: 1 VSGIKVQYKXNDSAPGDNQ.....EQDWTKYTVYVNGVLWGTG 150

Sequence: 1 VSGIKVQYKXNDSAPGDNQ.....EQDWTKYTVYVNGVLWGTG 150

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /Geneseq_032802:.*
2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	370	45.9	1352	22	AA63962
2	367	45.5	551	18	AAW18790
3	366	45.4	1350	22	AA63963
4	365	45.3	782	12	AA15625
5	361.5	44.9	700	12	AA13427
6	359.5	44.6	476	21	AA54123
7	359.5	44.6	493	20	AA28850
8	359.5	44.6	493	20	AA43218
9	358.5	44.5	167	17	AA95080
10	357.5	44.4	531	18	AA15238
11	357.5	44.4	1853	19	AAW43108

12	355	44.0	499	14	AA42122
13	340	42.2	616	20	AA13494
14	339	42.1	1751	20	AA13493
15	336	41.7	1426	20	AA13492
16	329	28.4	162	20	AA90077
17	228	28.3	162	15	AA63634
18	228	28.3	163	22	AAE05745
19	228	28.3	341	22	AAE05747
20	228	28.3	341	22	AAE05749
21	228	28.3	382	20	AA39952
22	228	28.3	428	22	AAE05748
23	227	28.2	190	22	AAE05746
24	226	28.0	154	20	AAW90081
25	226	28.0	156	20	AAW90080
26	130	16.1	531	16	AAW01503
27	130	15.0	532	12	AA133229
28	121	15.0	256	22	AA81128
29	84.5	10.5	475	22	AA692298
30	83	10.3	1723	22	ABG24680
31	82.5	10.2	986	21	AA40440
32	82.5	10.2	1346	22	AAU04567
33	82.5	10.2	1346	22	AAU04581
34	82.5	10.2	1371	22	AAU04570
35	81.5	10.1	633	21	AA84965
36	81	10.0	646	22	AA85086
37	80	9.9	1992	17	AAW04505
38	80	9.9	1992	22	AA691337
39	80	9.9	1992	22	AA691337
40	80	9.9	2047	22	AA691334
41	79.5	9.9	585	21	AA49311
42	79.5	9.9	666	21	AA49310
43	79.5	9.9	724	21	AA49309
44	79	9.8	237	11	AA805803
45	79	9.8	799	22	AA49462

ALIGNMENTS

RESULT 1	AA63962	standard; Protein; 1352 AA.
XX	AA63962:	
AC	XX	
XX	XX	
DT	29-OCT-2001	(first entry)
XX	XX	
DE	XX	Amino acid sequence of xyloglucanase enzyme.
XX	XX	xyloglucanase; family 44; glycosyl hydrolase; detergent;
KW	KW	cellulosic fiber; textile scouring.
XX	XX	
OS	XX	Paenibacillus polymyxa.
XX	XX	
PN	XX	W0200162903-A1.
PD	XX	30-AUG-2001.
PF	XX	21-FEB-2001; 2001WO-DK00116.
PR	XX	24-FEB-2000; 2000DK-0000291.
PA	XX	(NOVO) NOVOZYMES AS.
PI	XX	Schnorr K, Jorgensen PL, Schuelein M;
DR	XX	WPI; 2001-522819/57.
XX	XX	N-PSDB; AAH75059.
PT	XX	New xyloglucanase enzyme belonging to glycosyl hydrolases family,
PT	XX	useful for detergent compositions, and textile or cellulose fiber
XX	XX	processing industries

NK-1 cellulase. B
Truncated cellulases
Truncated cellulases
Truncated cellulases
C. cellulovorans C
Cellulose binding
Clostridium cellul
Clostridium cellul
Clostridium cellul
Gauzia luciferrase
Clostridium cellul
Clostridium cellul
C. cellulovorans C
C. cellulovorans C
60 kD endoglucanase
Endoglucanase enco
C17F2 Ospa constru
C glutamic prote
Novel human diagn
Human brain deriva
Human G-protein co
Human G-protein co
Human G-protein co
Amino acid sequenc
Bacillus lichenifo
Moraxella 200 kDa
M. catarrhalis str
M. catarrhalis str
M. catarrhalis str
Arabidopsis thalia
Arabidopsis thalia
Heat-stable endo-b
Propionibacterium

PS Example 3; Page 81-85; 97pp; English.
XX
CC The present sequence represents a xyloglucanase of the invention. The
CC specification describes a xyloglucanase enzyme belonging to family 44
CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
CC detergent compositions and prevents binding of certain soils to the
CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
CC room temperature and has a half life of more than 50 days when incubated
CC in a full formulated liquid detergent at 30 plus degrees Celsius. The
CC enzyme is used in detergent compositions, textile industry for improving
CC the properties of cellulosic fibers, yarn, and woven or non-woven
CC fabrics, preferably in textile scouring process, and in cellulose fiber
CC processing industry for retting of fibers e.g. hemp, jute, flax, and
CC linen.
XX
SQ Sequence 1352 AA;

Query Match 45.9%; Score 370; DB 22; Length 1352;
Best Local Similarity 46.0%; Pred. No. 9,8e-28;
Matches 69; Conservative 30; Mismatches 47; Indels 4; Gaps 3;

QY 4 GLKVVQKNNDSAPGDNQIKPGLQVNTGSSVDLSTVTVRWFTTRDGSSTLYVNCMDMA 63
DB 1200 glllyrtdtkvndhlnpfgivnkgfslpineklrlyttldgdtreg-tfncdyat 1258
QY 64 MCGGNIRASFGSVNPAATPYADTYIQLSFT--GGLAAGSGTGEIQNRVNSDMSNFTETN 121
DB 1259 lscskngklvmeakaatgadylevsfnsdagvlapgsctgdlgrlthktdwsnynead 1318
QY 122 DYSY-GTNTTFODWTKTVTYVNGVLVWGT 150
DB 1319 dlysk9tqlctadhpkvltlmgvltwgt 1348

RESULT 2

AAW18790
ID AAW18790 standard; protein; 551 AA.

AC AAW18790;

DT 18-NOV-1997 (first entry)

DE Corrected Bacillus lautus (NCIMB 40250) endoglucanase Endo 3A.

XX Endoglucanase; Endo 3A; formation; localised; variation;

KW colour density; surface; dye; fabric; family 5; cellulose;

KW hydrolysatoin; p-nitrophenyl-beta-1,4-cellobioside; stone wash;

XX blue jeans; back staining.

OS Bacillus lautus.

PN WO9709410-A1.

PD 13-MAR-1997.

PF 03-SEP-1996; 96WO-DK00364.

PR 08-SEP-1995; 95DK-0000993.

PA (NOVO) NOVO-NORDISK AS.

PI Flich M, Onishi M, Schulein M, Toft AH;

DR WPI; 1997-192888/17.

XX Localised variation of colour density in the surface of a dyed
PT cellulosic fabric - uses cellulase compen. able to hydrolyse
PT p-nitrophenyl -beta-1,4-cellobioside
XX
PS Disclosure; Pages 15-17; 23pp; English.
XX

CC The present sequence is the corrected version of the incorrect
CC Bacillus lautus (NCIMB 40250) endoglucanase Endo 3A described in
CC WO9110732. Endo 3 can be used in novel method of forming localised
CC colour density variation on the surface of a dyed cellulosic
CC fabric. The method comprises agitating the fabric in an aqueous
CC medium (pH 6.5 to 9.0) containing a family 5 cellulase,
CC e.g. the present sequence, which can hydrolyse p-nitrophenyl-beta
CC -1,4-cellobioside, or a family 7 cellulase, and a mechanical
CC abrading agent or cellulose having abrading activity. Each
CC cellulase displays 30 % or more of its maximum activity at pH 7.
CC The process is useful to provide a stone washed look to blue jeans
CC without back staining.
XX
SQ Sequence 551 AA;

Query Match 45.5%; Score 367; DB 18; Length 551;
Best Local Similarity 47.0%; Pred. No. 6,1e-28;
Matches 71; Conservative 31; Mismatches 43; Indels 6; Gaps 4;

QY 2 SGLKVVQKNNDSAPGDNQIKPGLQVNTGSSVDLSTVTVRWFTTRDGSSTLYVNCMD 61
DB 402 tgnlvvykvgtlsatdnqmkpsfnknngtltpvnlsglklrlyftkd-gradmasafdw 460
QY 62 AAMGCGNIRASFGSVNPAATPYADTYIQLSFT--GGLAAGSGTGEIQNRVNSDMSNFT 119
DB 461 aqiganvsaat--anfigsnldlyvelsfasgsalpaq9glgdlqirmyktdwsnlne 518
QY 120 TNDYSY-GTNTTFODWTKTVTYVNGVLVWGT 149
DB 519 andysdgaktayadwnrvltlmgngtlwgt 549

RESULT 3

AAG63963
ID AAG63963 standard; protein; 1350 AA.

AC AAG63963;

DT 29-OCT-2001 (first entry)

DE Amino acid sequence of xyloglucanase enzyme.

XX xyloglucanase; family 44; glycosyl hydrolase; detergent;

KW cellulosic fiber; textile scouring.

XX Paenibacillus polymyxa.

OS Paenibacillus polymyxa.

FT Key location/Qualifiers

PN WO200162903-A1.

PD 30-AUG-2001.

PF 21-FEB-2001; 2001WO-DK00116.

PR 24-FEB-2000; 2000DK-0000291.

PA (NOVO) NOVOZYMES AS.

PI Schnorr K, Jorgensen PL, Schulein M;

DR WPI; 2001-522819/57.

DR N-PSDB; AAH75060.

XX New xyloglucanase enzyme belonging to glycosyl hydrolases family,
PT useful for detergent compositions, and textile or cellulose fiber
PT processing industries
XX
PS Disclosure; Page 86-89; 97pp; English.
XX


```

PD 09-MAY-1996.
XX
XX 26-OCT-1995; 95WO-US13813.
XX
XX 27-OCT-1994; 94IL-0111415.
XX
XX (TECR ) TECHNION RES & DEV FOUND LTD.
XX (UYRA-) UNITA RAMOT APPLIED RES & IND DEV LTD.
XX (YEDA ) YEDA RES & DEV CO LTD.
XX
XX Bayer EA, Lamed R, Morag E, Shoham Y, Wilchek M;
XX
XX WPI; 1996-239453/24.
XX
XX Modified cellulose-binding domain mols. - having attached hapten,
XX particle, biotin, to provide an affinity system for, e.g. sepsis,
XX assays, reactors, delivery etc.
XX
XX Claim 12; Page 34-35; 53pp; English.
XX
XX A modified cellulose binding domain (CBD) or fraction of it, linked
XX to a hapten moiety via one or more Cys or Lys residues and complexes
XX comprising the biotinylated CBD and a biotin-binding molecule
XX selected from modified or unmodified avidin or streptavidin or an
XX anti-biotin antibody, can be used in affinity chromatography, cell
XX separation, cell, protein and enzyme immobilisation, selective
XX removal of biological materials, multienzyme reactors, signal
XX immunoassays and drug (e.g. pesticide) delivery. The CBD molecule
XX can be modified with the hapten without affecting its high affinity
XX for cellulose, or chitin.
XX
XX Sequence 167 AA;
XX
XX Query Match 44.5%; Score 358.5; DB 17; Length 167;
XX Best Local Similarity 44.0%; Pred. No. 9.1e-28;
XX Matches 70; Conservative 34; Mismatches 44; Indels 11; Gaps 3;
XX
XX 1 VSGGLKVOYKKNDSAPGDNQIKRGQLVNTGSSVDLSYTVRYWFTTRDGGSSSTLYVNCND 60
XX ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
XX 5 vsnqlkvelynspdslnspqfktvntgssaldsklrltyrvtyvqgkqdtfw-cd 63
XX
XX 61 WAAAM-----GCGNIRASFGSVNPATPTADTYLQLSFTGTTLAAGSGTEIQRVVK 111
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
XX 64 haaligsngsyngltsnvkgtlvkmsstnnadtylelftgltlepgahv-qiggrfak 122
XX
XX 112 SDMSNFTETNDYSYGTNTFTFDQMTKYVYVNGVLWGTE 150
XX : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 123 ndwsnytgndysfkssqfvewdqvtaylngvltwvge 161
XX
XX
XX RESULT 10
XX AAW15238
XX ID AAW15238 standard; protein; 531 AA.
XX
XX AAW15238;
XX
XX 28-JAN-1998 (first entry)
XX
XX Scaffoldin protein from Clostridium thermocellum.
XX
XX Dockerin; Celd; Cipa; scaffoldin; cellulose binding domain;
XX Chromatographic separation; soluble substrate modification; CBD;
XX multi-enzyme delivery system; animal feed; paper production;
XX plant protection; pest control.
XX
XX Clostridium thermocellum.
XX
XX Key Location/Qualifiers
XX FH 1..153
XX FT /label= internal_repeat_element_1
XX FT 154..306
XX FT /label= internal_repeat_element_2
XX

```

```

FT Domain 239..531
FT /label= cellulose-binding
FT /note= "only 60% of the CBD is present"
XX
XX WO9714789-A2.
XX
XX 24-APR-1997.
XX
XX 16-OCT-1996; 96WO-US16485.
XX
XX 17-NOV-1995; 95US-0559968.
XX 17-OCT-1995; 95US-0005701.
XX
XX (GENV ) GENENCOR INT INC.
XX
XX Bott RR, Clarkson KA, Fowler T, Iiu C, Ward M, Xia H;
XX
XX WPI; 1997-245106/22.
XX
XX Composition with enzymes non-covalently bound to a peptide backbone
XX - used as a multi-enzyme delivery system, e.g. in food processing,
XX textiles and pest control
XX
XX Claim 10; Fig 6; 29pp; English.
XX
XX This protein sequence represents a scaffoldin based on the Cipa protein
XX of Clostridium thermocellum. The scaffoldin protein is used in a novel
XX composition that comprises at least 2 enzymes non-covalently bound to a
XX peptide backbone (i.e. present sequence). The scaffoldin comprises a
XX number of internal repeating units and at least one cellulose binding
XX domain (CBD). The CBD may be altered to modify its affinity for
XX cellulose, which may be desirable where cellulose binding would be
XX disadvantageous. The enzyme is bound to a repeating element of the
XX region of the enzyme, which binds to a repeating element of the
XX scaffoldin. The dockerin is preferably Cels (AAW15237) or Celd
XX (AAW15236). The composition can be used in reducing allergenicity,
XX producing synergistic effects, and facilitating selective modification
XX of substrate. By taking advantage of the cellulose binding domain of the
XX complex, the complex could be immobilised for use in chromatographic
XX separations or for soluble substrate modification. By adding the
XX scaffoldin domain, it is possible to recover enzymes, or to quantify
XX the amount of an enzyme in a solution. The composition could also be
XX used in a multi-enzyme delivery system which could be used in the food
XX industry, in food processing, animal feed, textiles, bioconversion,
XX pulp and paper production, plant protection and pest control, as a wood
XX preservative, topical lotions, and biomass conversions.
XX
XX Sequence 531 AA;
XX
XX Query Match 44.4%; Score 357.5; DB 18; Length 531;
XX Best Local Similarity 44.0%; Pred. No. 5.1e-27;
XX Matches 70; Conservative 34; Mismatches 44; Indels 11; Gaps 3;
XX
XX 1 VSGGLKVOYKKNDSAPGDNQIKRGQLVNTGSSVDLSYTVRYWFTTRDGGSSSTLYVNCND 60
XX ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
XX 312 vsnqlkvelynspdslnspqfktvntgssaldsklrltyrvtyvqgkqdtfw-cd 370
XX
XX 61 WAAAM-----GCGNIRASFGSVNPATPTADTYLQLSFTGTTLAAGSGTEIQRVVK 111
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
XX 371 haaligsngsyngltsnvkgtlvkmsstnnadtylelftgltlepgahv-qiggrfak 429
XX
XX 112 SDMSNFTETNDYSYGTNTFTFDQMTKYVYVNGVLWGTE 150
XX : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 430 ndwsnytgndysfkssqfvewdqvtaylngvltwvge 468
XX
XX
XX RESULT 11
XX AAW43108
XX ID AAW43108 standard; protein; 1853 AA.
XX
XX AAW43108;
XX

```

DT 16-OCT-1998 (first entry)
 XX
 DE C. thermocellum cellulosome integrating protein.
 XX
 KM Multimer; enzyme; complex; protein-protein interaction; dockerin domain;
 KM cohesin domain; catalytic subunit; scaffold subunit; SdbA; synergistic;
 KM cellulosome integrating protein; scaffoldin dockerin binding protein.
 XX
 OS Clostridium thermocellum.
 XX
 PN FR2748479-A1.
 XX
 PD 14-NOV-1997.
 XX
 PF 10-MAY-1996; 96FR-0005854.
 XX
 PR 10-MAY-1996; 96FR-0005854.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Beguin P, Leibovitz E;
 XX
 DR MPI: 1998-011569/02.
 DR N-PSDB: AAT86625.
 XX
 PT Cellulase proteins with cohesin or dockerin type II domains - useful
 PT for potentiating the activity of multiprotein enzyme complexes
 XX
 PS Disclosure; Page 47; 60pp; French.

CC Multimeric protein, especially enzymatic, complexes are held together
 CC by protein-protein interactions between domains designated dockers
 CC and cohesins, which are found on the catalytic and scaffold subunits
 CC respectively. An example of such a complex is the cellulose degrading
 CC protein complex from Clostridium thermocellum, known as the cellulosome.
 CC This complex comprises around 15 proteins including endoglucanases,
 CC cellobiohydrolases, hemicellulases, e.g. xylanases or lichenases, which
 CC interact with a central "scaffold" protein designated the cellulosome
 CC integrating protein (Cip; see AAW43108). The catalytic subunits
 CC interact with the Cip subunit via conserved 23 amino acid dockerin
 CC domains. Cip has been shown to contain 9 copies of a cohesin domain.
 CC The invention relates to the isolation of proteins binding to a novel
 CC dockerin type domain found in the C-terminal portion of Cip. The new
 CC domain is designated a type II dockerin domain (as compared to the type
 CC I domain found on the catalytic subunits of the cellulosome). The type
 CC II dockerin domain has some sequence similarity to the type I dockers
 CC but is unable to bind type I cohesin domains.
 CC The sequence presented here is the cellulosome integrating protein
 CC (Cip-A) to which the proteins SdbA, OIPB and ORF2P (AAW43105-W43107) can
 CC bind at the C-terminal dockerin type II domain (location not given in
 CC specification). The novel type II dockerin and cohesin domains can be
 CC used in complexes, especially enzyme complexes, to potentiate their
 CC catalytic actions in a synergistic manner.

XX
 XX Sequence 1853 AA:

Query Match 44.4%; Score 357.5; DB 19; Length 1853;
 Best Local Similarity 44.0%; Pred. No. 2.5e-26;
 Matches 70; Conservative 34; Mismatches 44; Indels 11; Gaps 3;

OY 1 VSGGLKVOYKKNDSAPGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSTLYVNCMD 60
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||
 DB 365 vsgnlkvefynspadtnslpqfvtntgsaidskrltlyyvtvdgkqqtew-cd 423
 OY 61 WAAAM-----GCCNIRASFGSVNPAFPTADTYLQLSFTGCTLAAGSGTGEIONRVNK 111
 DB || : : : : : || : : : : : || : : : : : || : : : : : || : : : : : ||
 DB 424 haaligsngsyngltanvkvftkmsstlnadlylelftggtlepgahv-qiggrfak 482
 OY 112 SDMSNFTETNDYSYGTNTTFODTKVTYVYVNGVLWNGTE 150
 DB ||||| : : : : : || : : : : : || : : : : : || : : : : : || : : : : : ||
 DB 483 ndksnytgandysfkasagfvekdqvtaylmgvlgke 521

RESULT 12
 ID AAR42122 standard; Protein; 499 AA.
 XX
 AC AAR42122;
 XX
 DT 27-APR-1994 (first entry)
 XX
 DE NK-1 cellulase.
 XX
 KM Cellulase; pH dependence; mutation.
 XX
 OS Bacillus N4.
 XX
 PN JP05236969-A.
 XX
 PD 17-SEP-1993.
 XX
 PF 28-FEB-1992; 92JP-0075883.
 XX
 PR 28-FEB-1992; 92JP-0075883.
 XX
 PA (BEPF/) BEPPU T.
 XX
 DR MPI: 1993-330585/42.
 DR N-PSDB: AAQ49820.
 XX
 PT Changing the pH-dependence of cellulase enzymatic activity - by
 PT changing base sequence of cellulose-producing gene of Bacillus
 PT microbe to base sequence coding asparagine and serine residues at
 PT specified aminoacid sites
 XX
 PS Disclosure; Fig 1; 9pp; Japanese.

CC Sequences (AAQ49818-19) consist of two synthetic fragments which are
 CC used to induce a mutation within the cellulase NK-1 gene. The gene
 CC is shown in sequence (AAQ49820). The modified NK-1 gene shows a
 CC change in pH dependence.

XX
 XX Sequence 499 AA:

Query Match 44.0%; Score 355; DB 14; Length 499;
 Best Local Similarity 43.9%; Pred. No. 8.3e-27;
 Matches 65; Conservative 31; Mismatches 48; Indels 4; Gaps 2;

OY 4 GLKVOYKKNDSAPGDNOIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSTLYVNCMDA 62
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||
 DB 353 glavgykagdgynsqirpqlhknngnatvdlkdvltarywnaknkgqn---fdcdya 409
 OY 63 AMCGNIRASFGSVNPAFPTADTYLQLSFTGCTLAAGSGTGEIONRVNKSDMSNFTETND 122
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : ||
 DB 410 qlgcgnltkhfvclhpkpgadcytlelgtetlpsnscgnqlrlhnddwanyaqsgd 469
 OY 123 YSYGTNTTFODTKVTYVYVNGVLWNGTE 150
 DB || : : : : : || : : : : : || : : : : : || : : : : : || : : : : : ||
 DB 470 ysifgsntfkttkitlyhngkiliwte 497

RESULT 13
 ID AAY13494 standard; Protein; 616 AA.
 XX
 AC AAY13494;
 XX
 DT 30-JUL-1999 (first entry)
 XX
 DE Truncated cellulase Cel E3/B5.
 XX
 KM Cellulase; proteolase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
 KM Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
 KM cotton-containing fabric; stonewashing.

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:24:03 ; Search time 72.78 Seconds
(without alignments)
198,041 Million cell updates/sec

Title: US-09-917-384-5
Perfect score: 806
Sequence: 1 VSGIKVQYKNNDSAPGDNQ.....FQDWTKVTYVYVNGVLWGT 150

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	380	47.1	145	2	A41897 cellulase homolog
2	364	45.2	508	2	G69593 cellulase (EC 3.2.1)
3	362	44.9	499	2	JN0111 cellulase (EC 3.2.1)
4	361.5	44.9	700	2	B41897 cellulase (EC 3.2.1)
5	357.5	44.4	1854	2	S36859 cipa protein - C10
6	357	44.3	508	2	A26874 cellulase (EC 3.2.1)
7	355	44.0	499	2	A27198 cellulase (EC 3.2.1)
8	348.5	43.2	1711	2	T31337 1,4-beta-glucanase
9	341	42.3	1779	2	T31085 xylanase - Caldice
10	336	41.7	1331	2	A48954 mannan endo-1,4-be
11	335.5	41.6	504	2	S54744 cellulase (EC 3.2.1)
12	335	41.6	915	2	A43802 cellulase (EC 3.2.1)
13	335	41.6	1039	2	S02711 cellulase (EC 3.2.1)
14	335	41.6	1742	2	T17120 cellulase (EC 3.2.1)
15	335	40.3	505	2	S39962 endoglucanase - Er
16	315	39.1	486	2	T40548 bifunctional cellu
17	286	35.5	986	2	S12021 thermolabile cellu
18	282.5	35.0	1483	2	C97012 probably celluloso
19	270	33.5	1162	2	T30433 scaffolding protei
20	265	32.9	586	2	PC6006 scaffolding protei
21	254.5	31.6	1230	2	S47404 cellulase 1,4-beta
22	253	31.4	879	2	A47704 endoglucanase I (E
23	228	28.3	1848	2	A44140 cellulose-binding
24	112	13.9	1045	2	A39199 endoglucanase B (E
25	94.5	11.7	1428	2	AC2224 endo-beta-1,3-1,4-
26	92.5	11.5	269	2	H95976 endo-beta-1,3-1,4-
27	90.5	11.2	698	2	D90771 hypothetical prote
28	90.5	11.2	698	2	H85633 hypothetical prote
29	90.5	11.2	725	2	JC1300 endo-beta-1,4-glu

30	89.5	11.1	636	2	JC5874 cellulase (EC 3.2.1)
31	88.5	11.0	698	2	F64839 yegh protein precu
32	87.5	10.9	319	2	T43040 hypothetical prote
33	87.5	10.9	712	2	F97012 hypothetical prote
34	87.5	10.9	2348	2	AD1841 hypothetical prote
35	85.5	10.6	1910	2	AF0394 probable adhesin h
36	85	10.5	268	2	S34804 exok protein - Rhl
37	85	10.5	582	2	F70675 probable PPE prote
38	84.5	10.5	5188	2	B85547 probable Rfx fam1
39	84.5	10.5	5291	2	F90696 hypothetical prote
40	84	10.4	781	2	T36143 probable secreted
41	82.5	10.2	618	2	T08685 hypothetical prote
42	82	10.2	1649	2	C86822 hypothetical prote
43	81.5	10.1	491	2	T20061 hypothetical prote
44	81.5	10.1	2204	2	A70524 probable PPE prote
45	81	10.0	238	1	S19012 licheninase (EC 3.

ALIGNMENTS

RESULT 1

A41897 cellulase homolog - Bacillus lautus (fragment)
C/Species: Bacillus lautus
C/Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 15-Oct-1999
C/Accession: A41897; S27498
R:Hansen, C. K.; Diderichsen, B.; Jorgensen, P. L.
J. Bacteriol. 174, 3522-3531, 1992
A/Title: cels from Bacillus lautus Pl236 encodes a novel cellulose-binding endo-beta-
A/Reference number: A41897; MUID:92276330
A/Accession: A41897
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1145<HAN>
A/Cross-references: EMBL:M76586; NID:g142661; PIDN:AAA22302.1; PID:g142662
A/Experimental source: PL236
A/Note: sequence extracted from NCBI backbone (NCBIP:104604)

Query Match 47.1% Score 380; DB 2; Length 145;
Best Local Similarity 50.0%; Pred. No. 3; Le-27;
Matches 73; Conservative 29; Mismatches 38; Indels 6; Gaps 4;
QY 7 VOYKNNDSAPGDNQIKPGLQVNTGSSVSDSTVRYMFRDGGSSITLVYNCDAAMGC 66
Db 1 LQYAAADTNADNDIKSFNKNNGTSANVDSTIKIRYFKDSSAANVGW-IMAGLGG 59
QY 67 GNIRASFGSVNPAFPTADYTLQLSFT--GGTLAAGSTGELIONRVNKSDDMSNFTETNDYS 124
Db 60 SNIGISFG--NHTGNSDITVELSFSSSEAGSIAGGSGGSETQLMKSMTDMENFNEANDYS 117
QY 125 Y-GTNTTFQDWTKVTYVYVNGVLWGT 149
Db 118 FDGTKTFADMDRVYVLYONGOIWGT 143
RESULT 2
G69593 cellulase (EC 3.2.1.4) bgIC precursor - Bacillus subtilis
N/Alternate names: endo-1,4-beta-glucanase
C/Species: Bacillus subtilis
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
C/Accession: G69593; A26114; I40353; S24239; S49103; I39803
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerston, P.T.; Ertlan, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holstappel, S.; Hosono, S.; Hullo, M
koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rigter, M.; Rivoletta, C.; Rocha, E.; Rochie B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon
A. Authors: Schlecht, S.; Schroeter, R.; Scroffone, F.; Sekiguchi, J.; Sekowska, A.; Serd
kechul, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Tojroni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipst, A.; Zumstein, E.; Yoshikawa, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A. Authors: Yoshikawa, H. F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A.Reference number: A69580; MUID:98044033
A:Accession: 669593
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-508 <KUN>
A:Cross-references: GB:299113; GB:AL009126; NID:g2634090; PIDN:CA613696.1; PID:e1183471
A:Experimental source: strain 168
R:McKay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moraneli, F.; Sel
Nucleic Acids Res. 14, 9159-9170, 1986
A>Title: Structure of a *Bacillus subtilis* endo-beta-1,4-glucanase gene.
A.Reference number: A26114; MUID:87066783
A:Accession: A26114
A:Molecule type: DNA
A:Residues: 10-508 <MAC>
A:Cross-references: GB:X04689; NID:g39823; PIDN:CAA28392.1; PID:g39824
A:Experimental source: strain PAP115
A>Note: part of this sequence, including the amino end of the mature form, was confirmed
R:Lindahl, V.; Aa, K.; Tronamo, A.
A>Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from *Bacillus subtilis* C
A.Reference number: I40353; MUID:95225656
A:Accession: I40353
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 10-291, 'N', 293-508 <LIN1>
A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA7429.1; PID:g39777
R:Lindahl, V.; Aa, K.
A>Title: Submitted to the EMBL Data Library, June 1992
A.Reference number: S24239
A:Accession: S24239
A:Molecule type: DNA
A:Residues: 10-291, 'N', 293-508 <LIN2>
A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA7429.1; PID:g39777
A:Experimental source: strain CK-2
R:Wolf, M.; Geczli, A.; Borliss, R.
A>Title: Submitted to the EMBL Data Library, December 1993
A:Description: Genes encoding beta glucan-hydrolyzing enzymes in *Bacillus subtilis*: cons
A.Reference number: S49103
A:Accession: S49103
A:Molecule type: DNA
A:Residues: 10-508 <WOL>
A:Cross-references: EMBL:Z29076; NID:g509266; PIDN:CAA82317.1; PID:g509267
R:Seo, Y.S.; Lee, Y.H.; Pek, U.H.; Kang, H.
A>Title: Analysis on the nucleotide sequence of the signal region of *Bacillus subtilis*
A.Reference number: I39803
A:Accession: I39803
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 10-13, 'V', 15, 'I', 17-18, 'V', 20-21, 'F', 23, 'V', 25-26, 'AT', 29-31, 'P', 33, 'PQ', 36
A:Cross-references: GB:M8634; NID:g142657; PIDN:AAA22300.1; PID:g142658
A:Experimental source: strain ATCC 6633
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal prote
C:Genetics:
A:Gene: bglC
A:Function:
A:Pathway: cellulose degradation
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
I:38/Domain: signal sequence #status predicted <IG>
I:39-508/Product: cellulase #status predicted <AT>

Query Match	45.2%	Score 364	DB 2	length 508
Best Local Similarity	44.6%	Pred. No. 3.7e-25		
Matches 66	Conservative 32	Mismatches 46	Indels 4	Gaps 2

OY 4 GLKVOYKNNDSPAPDNOKPKPELOLVNTGNSVDDJSTVAVRWVF--TPROGSSTLYVNCMDMA 62
 || : || : | ||| || : | : || : || : || : || : || : || : || : || : || : || :
 Db 362 GISVQYRAGDSNMSNQIRPOLQIKNNNGTIVVDLKDVTAARYAKKNGQN---PDCDYA 418
 OY 63 AMGCNNTIRASFGSVNPATPTADTYLQLSFTGGTLAAGSGTCGEIONRVKSDMSNFETEND 122
 || : || : | || : || : || : || : || : || : || : || : || : || : || : || : || :
 Db 419 QIGCGNVHTHKVTLLHKKPGKADFTYLELGFKKNGTILAPGASTGNIOLRHLNDMSNYVAOSGD 478
 OY 123 YSYGNTTFODMTKVTVYVNVGLWVGTE 150
 || : || : | || : | || : | || : | || : | || : | || : | || : | || : | || : | || :
 Db 479 YSFERSNTEFKTKTKITTYLDGSKLIWGTE 506

RESULT 3
 JN0111 cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain BSE616)
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: Bacillus subtilis
 C>Date: 17-Apr-1993 #sequence-revision 17-Apr-1993 #text-change 28-May-1999
 C:Accession: JN0111
 R:Park, S.H.; Kim, H.K.; Pack, M.Y.
 Agric. Biol. Chem. 55, 441-448, 1991
 A>Title: Characterization and structure of the cellulase gene of Bacillus subtilis BS
 A:Reference number: JN0111; MUID:91299280
 A:Accession: JN0111
 A:Molecule type: DNA
 A:Residues: 1-499 <PAR>
 A:Cross-references: GB:001057; NID:g216387; PIDN:BAA00859.1; PID:d1001323; PID:g216388
 C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
 C:Comment: The authors translated the codon ATA for residue 102 as Tyr
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 A:Pathway: cellulose degradation
 C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
 A:Title: Domain: signal sequence #status predicted <SIG>

```

Query March          44.9% Score 362; DB 2; Length 499;
Best Local Similarity 44.2%; Pred. No.5.Se-25;
Matches 65; Conservative 29; Mismatches 51; Indels 2; Gaps 1;

QY      4 GLKVOYKNNDSPADPNQIKPGILOVNTGSSVSVDLSTVVRWFTFRDGGSLTYVNCIDMAA 63
       1 : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :
Db      353 GISVQYRAGDSKMSNQRFPOLQIKNNNGTIVYDLKDVARWYNNAKKGNQGV--DCDYAQ 410
       1 : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :

QY      64 MGCCNIARSPESVNPATPTADTYIQLSFTTGTLAAGSGTEGIQNRVNKSQMSNFETENDY 123
       1 : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :
Db      411 LGCGNVYYKFETTLHKPKOGADTYLELGFKNGTGLAPGASTGIOURLHNDWMSNYASGDY 470
       1 : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :

QY      124 SYGTNTFPODWTKTVVYVYGVLWGTE 150
       1 : : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :
Db      471 SFFKSNFTTKTKITLYIDGKLINGTE 497
       1 : : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :

RESULT    4
B41897
cellulase (EC 3.2.1.4) - Bacillus laurus
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus laurus
C:Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text-change 15-Oct-1999
C:Accession: B41897; S27499
J.Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.
J. Bacteriol. 174, 3522-3531, 1992
A::Title: celsA from Bacillus laurus PU236 encodes a novel cellulose-binding endo-beta-
A::Reference number: A41897; MUID:92276330
A::Accession: B41897
A::Status: preliminary
A::Molecule type: DNA
A::Residues: 1-700 <HAN>
A::Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22303.1; PID:g142663
A::Experimental source: Fuz36
A::Note: Sequence extracted from NCBI backbone (NCBIP:104605)
A::Function:
```


Db 423 AQTASNVTFFVFLSSSVSGADYYLEIGRSGAGQLPKDGTGEIOMRPNKDMSTNYNO 482

0y 120 TNDYSY-GTNTTFODMTKVTYYVNGVLMGNE 150

Db 483 GNDMSWISQMTSYGENEKRVAYIDGVLMQOE 514

```

RESULT 11
S54744
cellulase (EC 3.2.1.4) CelVI precursor - Erwina carotovora (SCC 3193)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Erwina carotovora
A:Variety: SCC 3193
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999
C:Accession: S54744; S44996
R:Maee, A.; Helkinelmo, R.; Palva, E.T.
Mol. Gen. Genet. 247, 17-26, 1995
A:Title: Structure and regulation of the Erwina carotovora subspecies carotovora SCC3193
A:Reference number: S54744; MUID:95231512
A:Accession: S54744
A:Molecule type: DNA
A:Residues: 1-504 <MAE>
A:Cross-references: EMBL:X79241; NID:9493492; PIDN:CAA55823.1; PTD:9493493
C:Genetics:
A:Gene: celVI
C:Function:
A:Pathway: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
;1-32/Domains: signal sequence #status predicted <SIG>
;33-504/Product: cellulase #status predicted <MAT>

```

RESULT 12
A43802
cellulase (EC 3.2.1.4) / cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Caldocellum s
N/Alternate names: endo-1,4-beta-glucanase
C/Species: Caldocellum saccharolyticum
C/Date: 30-Jan-1993 #sequence_revision 30-Sep-1993 #text_change 10-Jul-1998
C/Accession: A43802
R: Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 56, 3117-3124, 1990
A/Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophile "
A/Accession number: A43802; MUID:91136262
A/Status: preliminary; nucleic acid sequence not shown; not compared with conceptual trans
A/Molecule type: DNA
A/Residues: 1-915 <SAU>
A/Cross-references: EMBL:X13602
C/Function:
A/Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A/Pathway: cellulose degradation
C/Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
C/Keywords: glycosidase; hydrolase; polysaccharide degradation
C/20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SVX>

Query Match	Similarity	41.6%	Score 335;	DB 2;	Length 915;
Best Local	Similarity	45.0%	Fred. No. 3e-22;		
Matches	Conservative	29;	Mismatches 50;	Indels 4;	Gaps 3.
Qy	3 GGLKQYKKNNDAPDNOIKPGLQVWNTGSSSVSDISTVYRWYFWRDGGSTLYVNCMA	62			
	: : : : :				
Db	349 GQIKVLVANKENSTNTTIRPMLKAYVNSGSSSIDLSRTYTRKTYVYDGRASAVS-DMA	407			
	: : : : :				
Qy	63 AMCGCNIRASFSGSVNPATPTADTYLQLSFTGT--LAAAGSTGEIOLRNKRNKSDMSNFTET	120			
	: : : : :				
Db	408 QIGASNVTFEFKFLSSVSAGADYILEIGRKSAGQLQPKDQIGELDIRNKKDMSWNYNG	467			
	: : : : :				
Qy	121 NDYSY-GTNTTFQDWTKYTVYVNGVLWMCTE	150			
	: : : : :				
Db	468 NDMSVLQASMTSYGENEKETVAYIDGLVWGOE	498			
	: : : : :				

RESULT 13
S02711
cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum
N:Alternate names: endo-1,4-beta-glucanase
N:Contains: cellulase (EC 3.2.1.4); cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91)
C:Species: Caldocellum saccharolyticum
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
C:Accession: S02711
R:Saul, D.J.; Williams, L.C.; Love, D.R.; Chamlay, L.W.; Bergquist, P.L.
Nucleic Acids Res. 17, 439, 1989
A:Title: Nucleotide sequence of a gene from Caldocellum saccharolyticum encoding for
A:Reference number: S02711; MUID:09098398
A:Accession: S02711
A:Molecule type: DNA
A:Residues: 1*1039 <SMD>
A:Cross-references: EMBL:X13602; NID:g40645; PIDN:CAA31936.1; PID:g40646
C:Genetics:
A:Gene: celB
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
C:Keywords: glycosidase; hydrolyase; polysaccharide degradation
F:1-78/Domain: signal sequence #status predicted <MG>
F:29-1039/Product: cellulase #status predicted <MG>
F:72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SV>

RESULT 14
T17120
cellulase (EC 3.2.1.-) precursor, thermoactive - *Caldocellum saccharolyticum*
C:Species: *Caldocellum saccharolyticum*
C:Date: 15-Oct-1999 #sequence:rev1510 15-Oct-1999 #text:change 17-Nov-2000
C:Accession: T17120; A43745
R:Te'o, V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
Title: CelA, another gene coding for a multidomain cellulase from the extreme thermophile *Caldocellum saccharolyticum*

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:38:34 ; Search time 38.44 Seconds
(without alignments)
151.091 Million cell updates/sec

Title: US-09-917-384-5

Perfect score: 806
Sequence: 1 VSGGLKVVQYKKNDSAPGDNQ.....FQDMTKVTVVNGVLVWGTE 150

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	380	47.1	145	1 YCEA_PAEIA	P29718 paenibacill
2	364	45.2	499	1 GUN2_BACSU	P10475 bacillus su
3	364	44.9	499	1 GUN3_BACSU	P23549 bacillus su
4	361.5	44.9	700	1 GUN4_PAEIA	P29719 paenibacill
5	359.5	44.6	772	1 CIPB_CLOTM	001866 clostridium
6	357.5	44.4	1853	1 CIPB_CLOTM	006851 clostridium
7	357	44.3	499	1 GUN1_BACSU	P07983 bacillus su
8	336	41.7	1331	1 MANB_CAUSA	P22533 caldocellum
9	335.5	41.6	504	1 GUNB_ERMCA	059395 erwinia car
10	335	41.6	1039	1 GUNB_ERMCA	P10474 c endogluc
11	335	41.6	1742	1 GUNB_ERMCA	P22534 caldocellum
12	325	40.3	503	1 GUN2_CLOSR	P47036 erwinia car
13	294.5	36.5	914	1 GUN2_CLOSR	P23659 clostridium
14	286	35.5	986	1 GUN1_CLOTM	002934 clostridium
15	253	31.4	879	1 GUN1_CLOTM	P38058 clostridium
16	228	28.3	1848	1 CBPA_CLOCL	059394 erwinia car
17	146.5	18.2	444	1 GUNB_ERMCA	P22525 cellulomona
18	112	13.5	1045	1 GUNB_ERMCA	P22525 cellulomona
19	112	13.5	269	1 EXOK_RHIME	P33693 rhizobium m
20	92.5	11.5	880	1 GUN4_THIEPU	P26221 thermomonos
21	90.5	11.2	725	1 GUN2_CLOSR	P28622 bacillus sp
22	89.5	11.1	636	1 GUN4_BACSU	P75882 escherichia
23	88.5	11.0	698	1 YMCB_ECOCI	P45797 paenibacill
24	81	10.0	238	1 GUNB_PAEIO	P11235 mumps virus
25	80.5	10.0	582	1 HEMA_MUMPM	030611 pseudomonas
26	80.5	10.0	1148	1 ICERK_PSESX	P81004 xenopus lae
27	80	9.9	282	1 ICERK_PSESX	P06620 pseudomonas
28	79.5	9.9	1200	1 GUNB_PAEIA	P23904 paenibacill
29	79	9.8	437	1 ICEN_PSEFL	047899 flavobacter
30	78.5	9.7	443	1 ELYS_FLAME	039163 wiseana iti
31	77.5	9.6	484	1 COAT_LIRV9	P55316 methanococ
32	77	9.6	215	1 FLAI_MEYVA	P55328 aspergillus
33	76.5	9.5	211	1 XYNI_ASAPM	

34	76.5	9.5	211	1 XYNI_ASPTU	P55331 aspergillus
35	76.5	9.5	582	1 HEMA_MUMPR	P10866 mumps virus
36	76.5	9.5	1196	1 ICEN_PSESX	O33479 pseudomonas
37	76	9.4	211	1 XYNI_ASAPM	P55329 aspergillus
38	76	9.4	282	1 FOR2_MEIGA	P82013 meleagris g
39	76	9.4	1157	1 GUN4_BACTO	O45733 bacillus th
40	76	9.4	1210	1 ICEN_PSEFL	P09815 pseudomonas
41	75	9.3	247	1 FLAI_THIEVO	P57719 thermoplasma
42	74.5	9.2	211	1 XYNI_ASAPM	P33557 aspergillus
43	74.5	9.2	523	1 PUR9_MYCTU	P71553 m blifunctio
44	74	9.2	498	1 VNUC_IAUSS	P18073 influenza a
45	74	9.2	517	1 GUN4_CLOLO	P54937 clostridium

ALIGNMENTS

RESULT 1
YCEA_PAEIA STANDARD; PRT; 145 AA.

AC P29718;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein in CELA 5' region (Fragment).
OS Paenibacillus laetus (Bacillus laetus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Paenibacillus.
OX NCBI_TaxID=1401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL236;
RX MEDLINE=92276330; PubMed=1592807;
RA Hansen C.K., Joergensen P.L., Diderichsen B.;
RT "celA from Bacillus laetus PL236 encodes a novel cellulose-binding
endo-beta-1,4-glucanase";
RT J. Bacteriol. 174:3522-3531(1992).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC EMBL: M76588; AAA2302.1;
DR PIR: A41897; A41897.
DR PIR: S27498; S27498.
DR HSRF: 006851; JNBC.
DR InterPro: IPR001956; CBD_3.
DR Pfam: PF00942; CBD_3; 1.
DR ProDom: PD001947; CBD_3; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 145 AA; 15782 MW; 95143A71B106AEB CRC64;

Query Match 47.1%; Score 380; DB 1; Length 145;
Best Local Similarity 50.0%; Pred. No. 5.3e-28;
Matches 73; Conservative 29; Mismatches 38; Indels 6; Gaps 4;

OY	7	VQYKNDNSAPGDNQIRKGLQVNTGSSVDSTVYRWFTRDGSSTLVYNCDAAMGC	66
OY	1	LQYADADTNAADNOIKPSFNKNGTSVADSLTKIRYFTDGSAAVNGW-IDNAOLGG	59
OY	67	GNIRASGSVNPAPPTADTYLQSLFT--GGTLAAGSGTEIONRNKNSWMTFTNYS	124
OY	60	SNIOISRE--NHGTNSDITVELSESGSIAGGOSGETOLRMSKIDMSFNEANDYS	117
OY	125	Y-GTWTFTQDMTKVTVVNGVLVWGTE	149
OY	118	FDGKTAFADMDRVVLVYONGVLVWGTE	143

```

RESULT 2
GUN2_BACSU STANDARD: PRT: 499 AA.
ID GUN2_BACSU
AC p10475;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Carboxymethyl-cellulase) (CMCase) (Cellulase).
GN BGIC OR GID OR EGIS.
OS Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
ON NCBI_TaxId=1423;
RX MEDLINE=87066783; PubMed=3024130;
RC STRAIN=PAP115;
RA Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
RT Moravell F., Seligy V.;
RL "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.";
RN Nucleic Acids Res. 14:9159-9170(1986).
RP SEQUENCE FROM N.A.
RC STRAIN=CK-2;
RX MEDLINE=95225656; PubMed=7710280;
RA Lindahl V., Aa K., Tronamo A.;
RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
RL Antonie Van Leeuwenhoek 66:327-332(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Rose M., Entian K.;
RT Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE OF 30-45.
RC STRAIN=CK-2;
RX MEDLINE=95225655; PubMed=7710279;
RA Aa K., Flengsrud R., Lindahl V., Tronamo A.;
RT "Characterization of production and enzyme properties of an
RN endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
RL compost soil.";
RN Antonie Van Leeuwenhoek 66:319-326(1994).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z29076; CAA82317.1; -
DR EMBL: X04689; CAA28392.1; -
DR EMBL: X67044; CAA47429.1; -
DR EMBL: Z73234; CAA97610.1; ALT_INIT.
DR EMBL: Z99113; CAB13696.1; ALT_INIT.
DR PIR: A26114; A26114.
DR HSSP: O85465; 1A3H.
DR Subtilist: BG10437; bg1c.
DR Interpro: IPR001956; CBD_3.
DR Interpro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.

```

```

KW Cellulose degradation: Hydrolase; Glycosidase; Signal;
KM Complete proteome.
FT SIGNAL 1 29
FT CHAIN 30 499
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
FT CONFLICT 283 283 S -> N (IN REF. 2).
SQ SEQUENCE 499 AA; 55287 MW; 8F735FF711B3EAE2 CRC64;

Query Match
Best Local Similarity 44.6%; Score 364; DB 1; Length 499;
Matches 66; Conservative 32; Mismatches 46; Indels 4; Gaps 2;

QY 4 GLKQYKNDSPAGDQNIKPGQLVNTGSSVDLSVTVRYWF-TRDGGSTLVYNCDA 62
DB 353 GISVQYRAGDGSWNSNQIRPOLIKNGMWTYDLKQVARYWKAKKNGQN--PDCDYA 409
QY 63 AMCCGNIRASFGSVNPAFTPATYTLQSLFTGGLAAGSGTGELQNRVKSMSNFTETMD 122
DB 410 QIGCGNVTHTKFTYLRKPGADPTYLELGFKNGLTAGASTGNQLRLHNDWMSNVAQSGD 469
QY 123 YSYGTNFTFQDMTKRYVYVNGVLWGTE 150
DB 470 YSFRKSNTRFKTKITLYDQGLIWGTE 497

RESULT 3
GUN3_BACSU STANDARD: PRT: 499 AA.
ID GUN3_BACSU
AC P23549;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Carboxymethyl-cellulase) (CMCase) (Cellulase).
GN BGIC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
ON NCBI_TaxId=1423;
RX MEDLINE=91299280; PubMed=1368694;
RA Park S.H., Kim H.K., Pack M.Y.;
RT "Characterization and structure of the cellulase gene of Bacillus
RN subtilis BSE616.";
RN Agric. Biol. Chem. 55:441-448(1991).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D01057; BAA00859.1; -
DR PIR: JN0111; JN0111.
DR HSSP: O85465; 1A3H.
DR Interpro: IPR001956; CBD_3.
DR Interpro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KM Cellulose degradation: Hydrolase; Glycosidase; Signal.

```

```
GUNA_PAEIA          STANDARD:      PRT;       700 AA.ID GUNA_PAEIA  
AC P29719;  
DT 01-APR-1993 (rel. 25, Created)  
DT 01-APR-1993 (rel. 25, Last sequence update)  
DT 15-JUL-1998 (rel. 36, Last annotation update)  
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
GN CELA.  
OS Paenibacillus laetus (Bacillus laetus).  
OC Bacteria; Firmicutes; Bacillales/Clostridiales group;  
OC Bacillus/Staphylococcus group; Paenibacillus.  
OX NCBI TaxID=1401.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PL236;  
RA MEDLINE=92276330; PubMed=1592807;  
RX Hansen C.K., Dierichsen B., Jørgensen P.L.;  
RL "Cella from Bacillus laetus PL236 encodes a novel cellulose-binding  
    endo-beta-1,4-glucohydrolase."  
RT J. Bacteriol. 174:3522-3531(1992).  
RT -1 CARBOHYDRATE ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
    linkages in cellulose.  
CC -1 PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL  
    PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.  
CC -1 SIMILARITY: BELONGS TO CELLULOSE FAMILY V (FAMILY 44 OF GLYCOSYL  
    HYDROLASES).
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sdb.ch/announce/ or send an email to license@isb-sdb.ch)

EMBL: M76588; AAA22303.1;

PIR: BA1897; BA1897.
PIR: S27499; S27499.
HSSP: Q06851; INBC.
InterPro: IPR001956; CBD_3.
Pfam: PF00942; CBD_3; 1.
ProDom: PD001947; CBD_3; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 700 ENDOGLUCANASE A.
FT DOMAIN 548 700 CELLULOSE-BINDING (BY SIMILARITY).

[illegible]

Tue Sep 3 13:42:04 2002

us-09-917-384-5.rsp

```

Matches 70: Conservative 34; Mismatches 44; Indels 11; Gaps 3;
OY 1 VSGGLKVOYKNDASAPGNOIKPGLQVYNTGSSVSTVRYWTFRDGSSSTLYVNC 60
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 365 VSGNLKVEEYNSNPSTNIPORFKVTNGTSALDLSKTLRYRYVDCGDKQTFW-CD 423
OY 61 WAM-----GCCNIRASFSGVNPATPTADYTLQSTFGGTLAAGSGTGEIONRYN 111
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 424 HAALIGNSGYNITSNVAGTFVKMSSPTNADTYLEISFTGGTLEPGAHY-QIOGRPAK 482
OY 112 SDMSNFTETNDYSGTNTTFOQWTKYTVYVNCVLWGTE 150
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 483 NDMNSNTGSDYSPKSAQGVEMDQVATYLVNGLWGKE 521

RESULT 7
GUNI_BACSU STANDARD: PRT: 499 AA.
ID GUNI_BACSU
AC P07983;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
GN BGLC OR GLD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_Taxid=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-DIG;
RC MEDLINE=87194581; PubMed=3106328;
RA Robson L.M., Chambliss G.H.;
RT "Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.";
RL J. Bacteriol. 169:2017-2025(1987).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
    linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
    HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M16185; AAA2496.1; ALT_INIT.
CC PIR: A26874; A26874.
CC HSSP: O85465; 1A3H.
CC InterPro: IPR001956; CBD_3.
CC InterPro: IPR001547; Glyco_hydro_F5.
CC Pfam: PF00942; CBD_3; 1.
CC Pfam: PF00150; cellulase; 1.
CC ProDom: PD001947; CBD_3; 1.
CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Signal.
KW SIGNAL.
FT CHAIN 1 499
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NICLOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
FT SEQUENCE 499 AA; 55187 MW; 339004EE95A65BE1 CRC64;

Query Match 44.3%; Score 357; DB 1; Length 499;
Best Local Similarity 44.6%; Pred. No. 2.8e-25;
Matches 66; Conservative 30; Mismatches 48; Indels 4; Gaps 2;
OY 4 GLKVOYKNDASAPGNOIKPGLQVYNTGSSVSTVRYWTFRDGSSSTLYVNC 62
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

```

```

Db 353 GVSQVYKAGCGRNSNQIRPOLHIKNKNATVDLKDVTARYWYNKKNQON---FDCCYA 409
OY 63 AMGCNIRASFSGVNPATPTADYTLQSTFGGTLAAGSGTGEIONRYNKSMSNFTETND 122
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 410 QMCCGNLTHKRFVTLHKPKQGADYTLGFKTGLISPEASGNTQLRLHDMDSNVVAQSGD 469
OY 123 VSGTNTTFOQWTKYTVYVNCVLWGTE 150
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 470 YSFOSNFTKTKITLKHQGLIWGTE 497

RESULT 8
MANB_CALSA STANDARD: PRT: 1331 AA.
ID MANB_CALSA
AC P22533;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta-mannanase/endoglucanase A precursor [includes: Mannan endo-1,4-
    beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-
    mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)].
GN MANA.
OS Caldectellum saccharolyticum (Caldectellulostripor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermotoga bacter group; Caldectellulostripor.
OX NCBI_Taxid=44001;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93119139; PubMed=1476429;
RX Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;
RT "The beta-mannanase from 'Caldectellum saccharolyticum' is part of a
    multidomain enzyme.";
RL Appl. Environ. Microbiol. 58:3864-3867(1992).
RN [2]
RP SEQUENCE OF 1-346 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Luehl E., Jaamat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
    gene coding for a beta-mannanase from the extremely thermophilic
    bacterium 'Caldectellum saccharolyticum'.";
RL Appl. Environ. Microbiol. 57:694-700(1991).
CC -1- FUNCTION: DEGRADATION OF HEMICELLULOSES. THE SECOND MOST ABUNDANT
    POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH
    MANNANASE AND ENDOGLUCANASE ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
    galactoglucomannans.
CC linkages in mannans, galactomannans, glucomannans, and
    galactoglucomannans.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
    linkages in cellulose.
CC -1- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES
    CELSIUS.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
    A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
    J (FAMILY 44 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L01257; AAA71887.1; -.
CC EMBL: M36063; AAA72861.1; -.
CC PIR: BA3745; BA3745.
CC PIR: A48954; A48954.
CC HSSP: Q06851; 1NRC.
CC InterPro: IPR001956; CBD_3.
CC InterPro: IPR001547; Glyco_hydro_F5.
CC Pfam: PF00942; CBD_3; 2.
CC Pfam: PF00150; cellulase; 1.
DR

```

DR PRODOM: PD001947; CBD_3; 2
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KM Hydrolyase; Glycosidase; Cellulose degradation; Signal;
 FT Multifunctional enzyme.
 FT SIGNAL 1 41
 FT CHAIN 42 1331
 FT DOMAIN 42 325
 FT DOMAIN 326 361
 FT DOMAIN 362 518
 FT DOMAIN 519 564
 FT DOMAIN 565 720
 FT DOMAIN 721 780
 FT DOMAIN 1331 1331
 FT ACT_SITE 162 162
 FT ACT_SITE 257 257
 FT CONFLICT 338 338
 FT CONFLICT 340 346
 FT SEQUENCE 1331 AA; 146892 MW; FFBGCA51BBB8DF0E0 CRC64;

Query Match 41.7%; Score 336; DB 1; Length 1331;
 Best Local Similarity 44.7%; Pred. No. 7.2e-23;
 Matches 68; Conservative 29; Mismatches 51; Indels 4; Gaps 3;

OY 2 SGGGLVQYKNNDSAPGDNQIKFGLQLVNTGSSVDLSTVTVYVWFTROGSSSTLYVNCW 61
 DB 364 SGLIVLVANKETSTNTTIRFWLVNVSSSIDLSRTLYWTVDERAQSALS-DW 422
 OY 62 AAMGCGNIRASFVSNPATPTADTYQLSTFTGT--LAAGSGTGEIONRVKSDMSNFT 119
 DB 423 AQIGASNTYFKFVKLSSVSAGDYLEIGFKSGAGLOPGCKRTGELQMFNDDMSNYNQ 482
 OY 120 TNDYST-GTNTFQDMTKVTVYVNGVLWGT 150
 DB 483 GNDMSWISQMTSGYGENEKVTAYIDGLVWGOE 514

RESULT 9
 GUNB_ERMCA STANDARD: PRT; 504 AA.
 ID GUNB_ERMCA
 AC 059395;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Endoglucanase VI precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase VI) (Cellulase VI)
 GN CELV1.
 OS Erynia carotovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Pectobacterium.
 OX NCBI_TaxID=554;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SCC3193;
 RX MEDLINE=95231512; PubMed=7715600;
 RA Mae A., Heikinhelmo R., Palva E.T.;
 RT "Structure and regulation of the Erynia carotovora subspecies
 RT carotovora SCC3193 cellulase gene celv1 and the role of cellulase in
 RT phytopathogenicity";
 RT Mol. Gen. Genet. 247:17-26(1995).
 RL -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).

CC EMBL; X79241; CAA55823.1; -.
 DR HSSP; 085465; 1A3H.
 DR InterPro; IPR001956; CBD_3.
 DR InterPro; IPR001547; Glyco_hydro_F5.
 DR Pfam; PF00942; CBD_3; 1.
 DR Pfam; PF00150; cellulase; 1.
 DR PRODOM: PD001947; CBD_3; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KM Cellulose degradation; Hydrolyase; Glycosidase; Signal;
 FT SIGNAL 1 31
 FT CHAIN 32 504
 FT DOMAIN 32 334
 FT DOMAIN 335 352
 FT DOMAIN 353 504
 FT ACT_SITE 168 168
 FT ACT_SITE 256 256
 FT SEQUENCE 504 AA; 54963 MW; 0D7EC747815655FA CRC64;

Query Match 41.6%; Score 335.5; DB 1; Length 504;
 Best Local Similarity 44.4%; Pred. No. 2.6e-23;
 Matches 68; Conservative 31; Mismatches 47; Indels 7; Gaps 4;

OY 2 SGGGLVQYKNNDSAPGDNQIKFGLQLVNTGSSVDLSTVTVYVWFTROG--GSSTLYVNC 59
 DB 354 TGDIVQYKNNVSNPDDAIRAFNFKNTGSPRKISDQVYVYFHDGKPGANLFV--- 410
 OY 60 DWAAGCGNIRASFVSNPATPTADTYQLSTFTGT--LAAGSGTGEIONRVKSDMSNFT 118
 DB 411 DWANVPNNIVYVSTGTPASTKARVYLVYASGSLQPGAEVGRVRIHAGDSNVN 470
 OY 119 ETNDYSYGTN-TTFQDMTKVTVYVNGVLWGT 150
 DB 471 ETNDYSYGPNTSYTNMWDKITVHDKGLTWGTE 503

RESULT 10
 GUNB_CALSA STANDARD: PRT; 1039 AA.
 ID GUNB_CALSA
 AC P10474;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Endoglucanase/exoglucanase B precursor [includes: Endoglucanase
 DE (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)
 DE (Cellulohydrolyase); Exoglucanase (EC 3.2.1.91) (Exocellulohydrolyase)]
 GN CELB.
 OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Thermoaerobacter group; Caldicellulosiruptor.
 OX NCBI_TaxID=44001;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89098398; PubMed=2789517;
 RA Saul D.J., Williams L.C., Love D.R., Chamley I.W., Bergquist P.I.;
 RT "Nucleotide sequence of a gene from Caldocellum saccharolyticum
 RT encoding for exoglucanase and endoglucanase activity";
 RT Nucleic Acids Res. 17:439-439(1989)
 RL -1- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL
 CC DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS
 CC AN ENDOGLUCANASE.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellobiose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION: BELONGS TO CELLULOSE FAMILY
 CC F (FAMILY 10 OF GLYCOSYL HYDROLASES).
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION: BELONGS TO CELLULOSE FAMILY
 CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC -----

QY 61 MAAGCGNIRASFGSVNPPATPDTYLLSFT--GGTLAAGSGTGEIQNRVNSDMSNFT 118
 DB 822 MASIGSSNVTGFEVKMDGATGTGADYLLLETFPPQAGTLEPGAST-EVQGRSKIDMDYT 880
 QY 119 ETNDYSYSG--TNTFQDMTKVTVYVNGVLWGTE 150
 DB 881 QTNDSYFNPTASSYVDNKITATVYISGNLYGIE 913

RESULT 14

GUNT_CLOSR STANDARD: PRT; 986 AA.

AC P23659;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Endoglucanase 2 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Thermococcus cellulase) (Avicelase I).
 GN CELZ.
 OS Clostridium stercorarium.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OC NCBI_TaxID=1510;

SEQUENCE FROM N.A., AND SEQUENCE OF 26-36 AND 475-486.

RA STRAIN-NCIB 11745;
 RA MEDLINE-91066838; PubMed-2250652;
 RA Jauris S., Ruecknagel K.P., Schwarz W.H., Kratzsch P.,
 RA Bronnenmeier K., Staudenbauer W.L.;
 RA "Sequence analysis of the Clostridium stercorarium celz gene encoding
 RA a thermococcal cellulase (Avicelase I): Identification of catalytic
 RA and cellulose-binding domains.";
 RT Mol. Genet. 223:258-267(1990).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SIMILARITY: CONTAINS MULTIPLE CELLULOSE-BINDING SITES.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSYL
 CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC EMBL: X55299; CAA39010.1; ALT-SEQ.

DR PIR: S12021; S12021.

DR HSSP: P26221; 1TF4.

DR InterPro: IPR001956; CBD_3.

DR InterPro: IPR001701; Glyco_hydro_9.

DR Pfam: PF00942; CBD_3; 2.

DR Pfam: PF00759; Glyco_hydro_9; 1.

DR ProDom: PD001947; CBD_3; 1.

DR ProSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.

DR ProSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.

DR Cellulose degradation: Hydrolyase; Glycosidase; Repeat; Signal.

KW SIGNAL

FT CHAIN 1 25 ENDOGLUCANASE Z.

FT REPEAT 26 986 DOMAIN C'.

FT REPEAT 498 641 DOMAIN B'.

FT REPEAT 651 738 DOMAIN A'.

FT REPEAT 744 831 DOMAIN B'.

FT REPEAT 854 986 DOMAIN C'.

FT DOMAIN 835 986 CELULOSE-BINDING (BY SIMILARITY).

FT ACT_SITE 400 400 BY SIMILARITY.

FT ACT_SITE 438 438 BY SIMILARITY.

FT ACT_SITE 447 447 BY SIMILARITY.

FT ACT_SITE 986 AA; 109512 MW; 180280982293690 CRC64;

Query Match 35.5%; Score 286; DB 1; Length 986;

Best Local Similarity 37.9%; Pred. No. 1,86-18; Indels 8; Gaps 3;
 Matches 58; Conservative 36; Mismatches 51;

QY 2 SGGKAYQVKNNDASACPDNDIKRGDQVNTGSSSVYLVYVWFTRDGSSTLVYNDW 61
 DB 837 TGVIOIOMFNNGTSDKTNQIMPRRLTNTGTPRLSDVKIRYITIDGKQDNFM--CDW 895
 QY 62 MAAGCGNIRASFGSVNPPATPDTYLLSFTGTLAAG---GSTGEIQNRVNSDMSNFT 117
 DB 896 SSVGSNNITGTFVKMAEPKRGADYLETGFTDG---AGYLPNOSIEVQNRKADMDY 952
 QY 118 ETNDYSYSGTNTTFTQDMTKVTVYVNGVLWGTE 150
 DB 953 IOTNDYSFSTNTSGSNDRLTYISGLVSGIE 985

RESULT 15

GUNT_CLOTM STANDARD: PRT; 879 AA.

AC Q02934;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Endoglucanase I precursor (EC 3.2.1.4) (EGI) (Endo-1,4-beta-glucanase)
 DE (Cellulase I).
 GN CELL.
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OC NCBI_TaxID=1515;

SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.

RA STRAIN-NCIB 10682;
 RA MEDLINE-93171873; PubMed-8436949;
 RA Hazlewood G.P., Davidson K., Laurie J.I., Huskisson N.S.,
 RA Gilbert H.J.;
 RA "Gene sequence and properties of Cell, a family E endoglucanase from
 RA Clostridium thermocellum.";
 RT J. Gen. Microbiol. 139:307-316(1993).
 CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-D-
 CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
 CC GLUCANS. PRINCIPALLY ACTIVE AGAINST BARLEY BETA-GLUCAN.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- PATHWAY: CELLULOSE DEGRADATION.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSYL
 CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC EMBL: L04735; AAA20892.1; -.

DR PIR: A47704; A47704.

DR HSSP: P26221; 1TF4.

DR InterPro: IPR001956; CBD_3.

DR InterPro: IPR001701; Glyco_hydro_9.

DR Pfam: PF00942; CBD_3; 2.

DR Pfam: PF00759; Glyco_hydro_9; 1.

DR ProDom: PD001947; CBD_3; 1.

DR ProSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.

DR ProSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.

DR Cellulose degradation: Hydrolyase; Glycosidase; Signal.

KW SIGNAL

FT CHAIN 1 55 ENDOGLUCANASE I.

FT REPEAT 56 879 CATALYTIC.

FT REPEAT 729 879 CELULOSE-BINDING (BY SIMILARITY).

FT ACT_SITE 448 448 BY SIMILARITY.

FT ACT_SITE 486 486 BY SIMILARITY.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: August 29, 2002, 16:37:45 ; Search time 125.36 Seconds

(Without alignments)
206.998 Million cell updates/sec

Title: US-09-917-384-5

Perfect score: 1 VSGGLKVOYKNNDSAPGDNQ.....FQDWTKVYVYNGVLWGTE 150

Sequence: 1 VSGGLKVOYKNNDSAPGDNQ.....FQDWTKVYVYNGVLWGTE 150

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

562222

Total number of hits satisfying chosen parameters:

562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP unc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *
15: SP rrvirus: *
16: SP bacteriophage: *
17: SP archaea: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	434	53.8	170	2	09RFX6	09rf6 caldbacilli
2	430	53.3	921	2	09L8L8	09l8l8 caldbacilli
3	426	52.9	930	2	09RFX5	09rf5 caldbacilli
4	371	46.0	997	2	092411	092411 caldbacilli
5	364	45.2	499	2	093706	093706 caldbacilli su
6	364	45.2	508	2	093L00	093l00 caldbacilli su
7	357	44.3	499	2	052731	052731 caldbacilli su
8	355	44.0	499	2	045532	045532 caldbacilli su
9	350	43.4	501	2	083012	083012 caldbacilli su
10	348.5	43.2	1711	2	P96311	P96311 anaerocellu
11	345	42.8	1000	2	024820	024820 thermophilu
12	343	42.6	1770	2	09X3P5	09x3p5 caldicellul
13	341	42.3	1779	2	052374	052374 caldicellul
14	340	42.2	996	2	09A0H0	09a0h0 caldicellul
15	339	42.1	1751	2	09A0G4	09a0g4 caldicellul
16	336	41.7	1426	2	09X3P6	09x3p6 caldicellul

17	316	39.2	261	2	09A0G7	09a0g7 caldicellul
18	315.5	39.1	1915	2	09RPL0	09rpl0 activibrio
19	315	39.1	486	2	045430	045430 bacilli sp
20	307.5	38.2	473	2	09RK75	09rk75 streptomyce
21	303.5	37.7	887	2	09L308	09l308 clostridium
22	284	35.2	2316	2	09FDJ9	09fdj9 bacteroides
23	282.5	35.0	1483	16	0977Y4	0977y4 clostridium
24	270	33.5	1162	2	082830	082830 clostridium
25	265	32.9	1546	2	045996	045996 clostridium
26	254.5	31.6	1230	2	059325	059325 clostridium
27	252.5	31.3	92	2	09L8L9	09l8l9 caldbacilli
28	206	25.6	321	2	046392	046392 clostridium
29	182.5	22.6	660	2	09L372	09l372 clostridium
30	129.5	16.1	202	2	09A0G6	09a0g6 caldicellul
31	123.5	15.3	197	2	09A0G1	09a0g1 caldicellul
32	98	12.2	715	2	065987	065987 clostridium
33	89.5	11.1	659	2	09R9H6	09r9h6 bacillus pu
34	87.5	10.9	319	3	P78669	P78669 achiosacch
35	87.5	10.9	712	16	097RK5	097rk5 clostridium
36	86	10.7	440	2	093L19	093l19 bacillus sp
37	85	10.5	428	3	012646	012646 neocallimas
38	85	10.5	582	16	P71868	P71868 mycobacteri
39	84.5	10.5	582	12	083345	083345 mumps virus
40	84	10.4	781	2	09S261	09s261 streptomyce
41	83.5	10.4	359	5	095WD5	095wd5 pristina le
42	83	10.3	936	2	093IE6	093ie6 ruminooccu
43	82.5	10.2	257	13	090536	090536 ginglymocto
44	82.5	10.2	438	10	09FHB7	09fhb7 arbidopsia
45	82.5	10.2	582	12	083347	083347 mumps virus

ALIGNMENTS

RESULT 1
ID 09RFX6 PRELIMINARY; PRT: 170 AA.
AC 09RFX6;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HYPOTHETICAL 18.5 KDA PROTEIN (FRAGMENT).
OS Caldbacillus cellulosovorans.
OC Bacteria; Firmicutes; Bacilli; Clostridium group;
OC Bacillus; Firmicutes; Bacilli; Clostridium group;
OC Caldbacillus cellulosovorans group; Caldbacillus.
OX NCBI_Taxid=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Suna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldbacillus cellulosovorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL: AF163837; AAF22273.1; -.
DR HSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR Pfam: PF00942; CBD_3; 1.
DR Prodom: PD001947; CBD_3; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 170 AA; 18493 MW; 7AC9D33E44E3A0B4 CRC64;

Query Match 53.8%; Score 434; DB 2; Length 170;
Best Local Similarity 52.3%; Pred. No. 1.2e-29;
Matches 79; Conservative 29; Mismatches 39; Indels 4; Gaps 3;

QY 3 GGLVQYKNNDSAPGDNQIKPQLQVMTGSSVDTSTVTVRWYFTROGSSLTLYVNCMA 62
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 20 GSLVQYRADPTNAGDNQKPKFRVYVNGTSSVPLSTIRYWTVD-GDPQVFNCDMA 78
QY 63 AMCGNIRASFGSVNPAITPRADTYLQLSFT--GGTLAAGSGTGEIONRVKNSDMSNFTET 120

```

DB 79 QVCGCNVNSGFYKLTSGRGADYIEITFTSGAGSIAAGASSGDIQVRINKNDWTNYNNA 138
OY 121 NDYSG-1NTTFQDWTKVTVYVNGVLWGTG 150
DB 139 NDYSYDPTKTSFADWNKRVLYLNGQLIMGVE 169

RESULT 2
O9L8L8 PRELIMINARY; PRT; 921 AA.
AC O9L8L8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BETA-1,4-XYLANASE XYNA PRECURSOR.
GN XYNA.
OS Caldicellulosus cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Alicyclobacillus group; Caldicellulosus.
OX NCBI_TaxID=74586;
RN [1]
RX MEDLINE-20519260; PubMed-11065373;
RA Sunna A., Gibbs M.D., Bergquist P.L.;
RT "A novel thermostable multidomain 1,4-beta-xyloxyase from
RT Caldicellulosus cellulovorans" and effect of its xyloxy-binding domain
RT on enzyme activity."
RL Microbiology 146:2947-2955(2000).
DR EMBL; AF200304; AAF61649.1; -.
DR HSSP; Q06851; 1NRC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR003305; CBD_6.
DR InterPro; IPR01000; Glyco_hydro.10.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF02018; CBD_6; 1.
DR Pfam; PF00331; Glyco_hydro.10.
DR PRINTS; PR00134; GHYDRASE10.
DR PRINTS; PR01217; PRICHEXTENS.
DR PRODOM; PD001947; CBD_3; 2.
KW Signal; Xylan degradation; Hydrolase; Glycosidase.
FT SIGNAL 1 33
FT CHAIN 34 921 POTENTIAL.
FT SEQUENCE 921 AA; 102380 MW; CSDDDIATF7567413 CRC64;

Query Match 53.3%; Score 430; DB 2; Length 921;
Best Local Similarity 53.3%; Pred. No. 1.8e-28;
Matches 80; Conservative 33; Mismatches 35; Indels 2; Gaps 2;

OY 2 SGLKVOYKNNDSAPGDNQIKPGLQIVNTGSSVDLSTVTVRYWFTRDGSSTLVYNCNM 61
OY 563 SGLTREVYVSGSATDQNMKPOLRIYVNTGSOAVPLTELKRYWYTKN-STQAEQYFCWM 621
OY 62 AAMGCNIRASFGSVNPAFTADTYIQLSFTGTLAAGSTGEIONRVKKSQMSNFTETN 121
DB 622 A0IGCSNIRAOVYSIAQPVSGADSYTELSTFGSGVPAAGNTGEIONRIHFTMMNMYNETD 681
OY 122 DYSY-GTNTTFQDWTKVTVYVNGVLWGTG 150
DB 682 DMSYNGQTTWGPSTRITLYLRNGVLWGTG 711

RESULT 3
O9RFXS PRELIMINARY; PRT; 930 AA.
AC O9RFXS;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MULTIDOMAIN BETA-1,4-MANNANASE PRECURSOR.
GN MANA.
OS Caldicellulosus cellulovorans.

```

```

OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Alicyclobacillus group; Caldicellulosus.
OX NCBI_TaxID=74586;
RN [1]
RX MEDLINE-20120520; PubMed-10653733;
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicellulosus cellulovorans and action of the recombinant enzyme on
RT kraft pulp."
RT Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL; AF163837; AAF22274.1; -.
DR HSSP; Q06851; 1NRC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR004302; Chitin_bind.3.
DR InterPro; IPR001547; Glyco_hydro.15.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF00942; CBD_3; 2.
DR Pfam; PF00150; cellulase.1.
DR Pfam; PF03067; Chitin_bind.3; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR PRODOM; PD001947; CBD_3; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Signal.
FT SIGNAL 1 33
FT CHAIN 34 930 MULTIDOMAIN BETA-1,4-MANNANASE.
FT SEQUENCE 930 AA; 101576 MW; 0086638D54D1A2CC CRC64;

Query Match 52.9%; Score 426; DB 2; Length 930;
Best Local Similarity 52.0%; Pred. No. 4.1e-28;
Matches 78; Conservative 35; Mismatches 35; Indels 2; Gaps 2;

OY 2 SGLKVOYKNNDSAPGDNQIKPGLQIVNTGSSVDLSTVTVRYWFTRDGSSTLVYNCNM 61
DB 287 SGLTREVYVSGSATDQNMKPOLRIYVNTGSOAVPLTELKRYWYTKN-STQAEQYFCWM 345
OY 62 AAMGCNIRASFGSVNPAFTADTYIQLSFTGTLAAGSTGEIONRVKKSQMSNFTETN 121
DB 346 A0IGCSNIRAOVYSIAQPVSGADSYTELSTFGSGVPAAGNTGEIONRIHFTMMNMYNETD 405
OY 122 DYSY-GTNTTFQDWTKVTVYVNGVLWGTG 150
DB 406 DMSYNGQTTWGPSTRITLYLRNGVLWGTG 435

RESULT 4
O9Z411 PRELIMINARY; PRT; 997 AA.
AC O9Z411;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLULASE PRECURSOR (EC 3.2.1.4).
GN CELB.
OS Bacillus sp. BP-23.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=89769;
RN [1]
RX MEDLINE-21129642; PubMed-11234960;
RA Pastor F.I.J., Puol J., Blanco A., Vidal T., Torres A.L., Diaz P.;
RT "Molecular cloning and characterization of a multidomain endoglucanase
RT from Paenibacillus sp BP-23: evaluation of its performance in pulp
RT refining."
RT Appl. Microbiol. Biotechnol. 55:61-68(2001).
CC -1-SIMILARITY: COMPAIR 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
DR EMBL; AJ13614; CAB38941.1; -.
DR HSSP; P26221; 1TFA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR003961; FN_III.

```

DR InterPro: IPR001101; Glyco_hydro_9.
 DR InterPro: IPR000566; Lipocalin_cytrFABP.
 DR Pfam: PF00942; CBD_3; 2.
 DR Pfam: PF00041; fn3; 2.
 DR Pfam: PF00759; Glyco_hydro_9; 1.
 DR ProDom: PD001947; CBD_3; 1.
 DR ProDom: PD00060; FN3; 2.
 DR SMART: SM00060; FN3; 2.
 DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
 DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
 DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
 DR GlycoSite: Hydrolyase; Signal.
 KW SIGNAL 1 34
 FT SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;

Query Match 46.0%; Score 371; DB 2; Length 997;
 Best Local Similarity 47.4%; Pred. No. 2.1e-23;
 Matches 72; Conservative 28; Mismatches 48; Indels 4; Gaps 3;

QY 2 SGLKVOYKNNDSAPGDNQIRPGLQLVNTGSSVDLSTVYRWFTRDGSSTLYNCDA 61
 ID 093JUV6 PRELIMINARY; PRT; 499 AA.
 AC 093JUV6
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE ALKALI TOLERABLE CELLULASE.
 GN CEL.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=7106;
 RA "J.M., Wang P., Liu X., Qu Y.;
 RT "The cloning and expression of a cellulase from Bacillus subtilis
 Y106.";
 RL EMBL: AF355629; AAK39540.1; -
 DR EMBL: AF355629; AAK39540.1; -
 SQ SEQUENCE 499 AA; 55232 MW; 587A687584B7E5BE CRC64;

RESULT 5
 ID 093JUV6 PRELIMINARY; PRT; 499 AA.
 AC 093JUV6
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE ALKALI TOLERABLE CELLULASE.
 GN CEL.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=7106;
 RA "J.M., Wang P., Liu X., Qu Y.;
 RT "The cloning and expression of a cellulase from Bacillus subtilis
 Y106.";
 RL EMBL: AF355629; AAK39540.1; -
 DR EMBL: AF355629; AAK39540.1; -
 SQ SEQUENCE 499 AA; 55232 MW; 587A687584B7E5BE CRC64;

Query Match 45.2%; Score 364; DB 2; Length 499;
 Best Local Similarity 44.6%; Pred. No. 3.8e-23;
 Matches 66; Conservative 32; Mismatches 46; Indels 4; Gaps 2;

QY 4 GLKVOYKNNDSAPGDNQIRPGLQLVNTGSSVDLSTVYRWFTRDGSSTLYNCDA 62
 ID 093JUV6 PRELIMINARY; PRT; 499 AA.
 AC 093JUV6
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE ALKALI TOLERABLE CELLULASE.
 GN CEL.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=7106;
 RA "J.M., Wang P., Liu X., Qu Y.;
 RT "The cloning and expression of a cellulase from Bacillus subtilis
 Y106.";
 RL EMBL: AF355629; AAK39540.1; -
 DR EMBL: AF355629; AAK39540.1; -
 SQ SEQUENCE 499 AA; 55232 MW; 587A687584B7E5BE CRC64;

RESULT 6
 ID 093JUV6 PRELIMINARY; PRT; 508 AA.
 AC 093JUV6
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4).
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CH21;
 RA "Yizdani C.J., Delgado O.D., Zavaia R., Mattiasson B.;
 RT "Bacillus subtilis CH21 endo-b-1,4-glucanase gene.";
 RL EMBL: AY044252; AAK94871.1; -
 DR EMBL: AY044252; AAK94871.1; -
 KW Hydrolyase; Glycosidase.
 FT SEQUENCE 508 AA; 56553 MW; A395E1C6F340F2 CRC64;

Query Match 45.2%; Score 364; DB 2; Length 508;
 Best Local Similarity 44.6%; Pred. No. 3.9e-23;
 Matches 66; Conservative 32; Mismatches 46; Indels 4; Gaps 2;

QY 4 GLKVOYKNNDSAPGDNQIRPGLQLVNTGSSVDLSTVYRWFTRDGSSTLYNCDA 62
 ID 093JUV6 PRELIMINARY; PRT; 499 AA.
 AC 093JUV6
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE ALKALI TOLERABLE CELLULASE.
 GN CEL.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=7106;
 RA "J.M., Wang P., Liu X., Qu Y.;
 RT "The cloning and expression of a cellulase from Bacillus subtilis
 Y106.";
 RL EMBL: AF355629; AAK39540.1; -
 DR EMBL: AF355629; AAK39540.1; -
 SQ SEQUENCE 499 AA; 55232 MW; 587A687584B7E5BE CRC64;

RESULT 7
 ID 052731 PRELIMINARY; PRT; 499 AA.
 AC 052731
 DT 01-JUN-1998 (TREMblrel. 06, Created)
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE ENDO-B-1,4-GLUCANASE (EC 3.2.1.4).
 GN CELS.
 OS Bacillus sp. 79-23.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=72363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=79-23;
 RA "Jung K.H., Chun Y.C., Lee J.-C., Kim J.H., Yoon K.-H.;
 RT "Cloning and expression of a Bacillus sp. 79-23 cellulase gene.";
 RL EMBL: AF045482; AAC02536.1; -
 DR EMBL: AF045482; AAC02536.1; -
 DR HSP: O85465; 1A3H.
 DR InterPro: IPR001956; Glyco_hydro_F5.
 DR InterPro: IPR001547; Glyco_hydro_F5.
 DR Pfam: PF00942; CBD_3; 1.
 DR ProDom: PD001947; CBD_3; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 KW Hydrolyase; Glycosidase.

DR InterPro: IPR001701; Glyco_hydro_9.
DR Pfam: PF00942; CBD_3; 3.
DR Pfam: PF02011; Glyco_hydro_48; 1.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR PRINTS: PR00844; GLHYDRLASE48.
DR ProDom: PD001947; CBD_3; 2.
DR ProDom: PD001903; Glyco_hydro_48; 1.
DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT NON_TER
SQ SEQUENCE 1711 AA; 189979 MW; E3E987CEB9CD0C21 CRC64;

Query Match 43.2%; Score 348.5; DB 2; Length 1711;
Best Local Similarity 46.1%; Pred. No. 3.5e-21;
Matches 72; Conservative 28; Mismatches 49; Indels 5; Gaps 4;

OY 1 VSGG-LKVOYKNDAPGDNOIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVYNC 59
DB 683 VAGGQIKVLYANKETNSTTIRPMLKVNTGSSSIDLSRVTIRYWTVDGDKAQSAIS- 741
OY 60 DNAMCCGNTIRASFGSVNPRATPTADTYLQLSFTGCT--LAAGSGTEIGONRVKSDMSNF 117
DB 742 DMAQIGASNVTFKFKVLLSSVSGADYLLLEIGFSGAGQLQAGDGTGEIQIRFKSDMSNY 801
OY 118 TETNDYSY-GTNTTFODMTKTVYVNGVLWGTE 150
DB 802 NQGNDSWMSQMTNVTGENKATVATIDGVLWGQE 835

RESULT 11
O24820 PRELIMINARY; PRT; 1000 AA.
AC 024820:
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BETA-GLUCANASE.
OS thermophilic anaerobe M10.
OC Bacteria.
OX NCBI_Taxid=67756;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M10;
RA Miyake K., Hattori K., Iijima S.;
RT "Characterization of a multi-domain cellulase from an extremely
thermophilic anaerobe strain M10."
RT Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB008029; BAA22939.1; -
DR HSSP: Q06851; INBC.
DR InterPro: IPR001589; Actinln_act_bind.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001000; Glyco_hydro_10.
DR InterPro: IPR001547; Glyco_hydro_F9.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS: PR00134; GLHYDRLASE10.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10_1; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
SQ SEQUENCE 1000 AA; 113265 MW; B9F659A56A752C6B CRC64;

Query Match 42.8%; Score 345; DB 2; Length 1000;
Best Local Similarity 46.1%; Pred. No. 3.5e-21;
Matches 70; Conservative 28; Mismatches 50; Indels 4; Gaps 3;
OY 2 SGLKVOYKNDAPGDNOIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVYNC 61
DB 372 SQGKVLVANKETNSTTIRPMLKVNTGSSSIDLSRVTIRYWTVDGDKAQSAIS-DW 430

OY 62 AAMCGNIRASFGSVNPRATPTADTYLQLSFTGCT--LAAGSGTEIGONRVKSDMSNFTE 119
DB 431 AQIGASNVTFKFKVLLSSVSGADYLLLEIGFSGAGQLQPGDGTGEIQIRFKSDMSNVNQ 490
OY 120 TNDYSY-GTNTTFODMTKTVYVNGVLWGTE 150
DB 491 GNDMSWISQMTNVTGENKATVATIDGVLWGQE 522

RESULT 12
O9X3P5 PRELIMINARY; PRT; 1770 AA.
ID O9X3P5:
AC O9X3P5:
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE XINA.
GN XINA.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoaerobacter group; Caldicellulosiruptor.
OX NCBI_Taxid=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1."
RT Curr. Microbiol. 40:333-340(2000).
DR EMBL: AF078737; AAD30363.1; -
DR HSSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR003305; CBD_6.
DR InterPro: IPR001000; Glyco_hydro_10.
DR InterPro: IPR002965; P_Rich_extensn.
DR Pfam: PF00942; CBD_3; 3.
DR Pfam: PF02018; CBD_6; 2.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS: PR00134; GLHYDRLASE10.
DR PRINTS: PR01217; PRICHEXTENSN.
DR ProDom: PD001947; CBD_3; 3.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10_1; 1.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10_1; 1.
SQ SEQUENCE 1770 AA; 193641 MW; 8BAF193704926C92 CRC64;

Query Match 42.6%; Score 343; DB 2; Length 1770;
Best Local Similarity 45.3%; Pred. No. 9.9e-21;
Matches 68; Conservative 29; Mismatches 49; Indels 4; Gaps 3;

OY 4 GLAVQYKNDAPGDNOIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVYNCMAA 63
DB 1104 GLAVLYKNNETSTASTSIRFWFKIVNGSSVDLSRVKIRYWTVDGDKPOSAY-CDMAQ 1162
OY 64 MCGNIRASFGSVNPRATPTADTYLQLSFTGCT--LAAGSGTEIGONRVKSDMSNFTE 121
DB 1163 IGASNVTFKFKVLLSSVSGADYLLLEIGFSGAGQLQPGDGTGEIQIRFKSDMSNVNQAD 1222
OY 122 DYSY-GTNTTFODMTKTVYVNGVLWGTE 150
DB 1223 DMSWISQMTNVTGENAKVTVYDGVLVWGQE 1252

RESULT 13
O52374 PRELIMINARY; PRT; 1779 AA.
ID O52374:
AC O52374:
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE FAMILY 10 XLANASE (EC 3.2.1.8).

GN XYNC.
OC Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=RT69B.1;
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RT "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp.
RT Rt69B.1.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF036924; AAB95326.1; -
DR HSSP: Q06851; INBC
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR003305; CBD_6.
DR InterPro: IPR001000; Glyco_hydro_10.
DR Pfam: PF00942; CBD_3; 3.
DR Pfam: PF02018; CBD_6; 2.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS: PR00134; GLYDRLASE10.
DR PRODOM: PD001947; CBD_3; 3.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
KW Xylan degradation; Hydrolase; Glycosidase.
SQ SEQUENCE 1779 AA; 194304 MW; CE5269B6806B5CED CRC64;

Query Match 42.3%; Score 341; DB 2; Length 1779;
Best Local Similarity 45.3%; Pred. No. 1.5e-20;
Matches 68; Conservative 27; Mismatches 51; Indels 4; Gaps 3;

OY 4 GLKVOYKNNDSAPGDNQIRPGIOLVNTGSSSYDLSTVYRVYFTRDGGSSSTLYVNCMDAA 63
DB 1113 GLKVLKNNETSASGSIRPWEKIVNGSSSYDLSTVYRVYFTRDGGSSSTLYVNCMDAA 1171
OY 64 MCGGNIRASFGSVNPAATPADIYLOLSTFGT--LAAGSGTEIONRVKSDMSNFTETN 121
DB 1172 IGASVNTFNFVKLSGVSADYLLVEGFSSGAGQLPGKDTDIDYVRFKNDMSVYNQAD 1231
OY 122 DYSY-GTNTTFQDMTKRYTVYVGVLVWGTE 150
DB 1232 DMSWQSMNTYGENAKVTLLYDGVLVWGQE 1261
RESULT 14
OYAOH0 PRELIMINARY; PRT; 996 AA.
AC OYAOH0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GLYCOSYL HYDROLASE 5 (FRAGMENT).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RA MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR HSSP: AF078038; AAK06388.1; -
DR HSSP: Q06851; INBC
DR InterPro: IPR002860; BNR.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR002965; P_Rich_extensn.
DR Pfam: PF02012; BNR; 9.
DR PRINTS: PR01217; PRICHEXTENS.
DR PRODOM: PD001947; CBD_3; 1.

KW Hydrolase.
FT NON_TER 996
SQ SEQUENCE 996 AA; 108275 MW; 3C72B6ED22F3C614 CRC64;

Query Match 42.3%; Score 340; DB 2; Length 996;
Best Local Similarity 45.3%; Pred. No. 9.2e-21;
Matches 68; Conservative 28; Mismatches 50; Indels 4; Gaps 3;

OY 4 GLKVOYKNNDSAPGDNQIRPGIOLVNTGSSSYDLSTVYRVYFTRDGGSSSTLYVNCMDAA 63
DB 844 GLKVLKNNETSASGSIRPWEKIVNGSSSYDLSTVYRVYFTRDGGSSSTLYVNCMDAA 902
OY 64 MCGGNIRASFGSVNPAATPADIYLOLSTFGT--LAAGSGTEIONRVKSDMSNFTETN 121
DB 903 IGASVNTFNFVKLSGVSADYLLVEGFSSGAGQLPGKDTDIDYVRFKNDMSVYNQAD 962
OY 122 DYSY-GTNTTFQDMTKRYTVYVGVLVWGTE 150
DB 963 DMSWQSMNTYGENAKVTLLYDGVLVWGQE 992

RESULT 15
OYAOH4 PRELIMINARY; PRT; 1751 AA.
AC OYAOH4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELE.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RA MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL: AF078042; AAK06394.1; -
DR HSSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001701; Glyco_hydro_9.
DR InterPro: IPR000566; Lipocin_cytfab.
DR InterPro: IPR002965; P_Rich_extensn.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR PRINTS: PR01217; PRICHEXTENS.
DR PRODOM: PD001947; CBD_3; 3.
DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
SQ SEQUENCE 1751 AA; 192176 MW; 60178CF3C00B95 CRC64;

Query Match 42.1%; Score 339; DB 2; Length 1751;
Best Local Similarity 44.7%; Pred. No. 2.1e-20;
Matches 67; Conservative 29; Mismatches 50; Indels 4; Gaps 3;

OY 4 GLKVOYKNNDSAPGDNQIRPGIOLVNTGSSSYDLSTVYRVYFTRDGGSSSTLYVNCMDAA 63
DB 678 GYKVLKNNETSASGSIRPWEKIVNGSSSYDLSTVYRVYFTRDGGSSSTLYVNCMDAA 736
OY 64 MCGGNIRASFGSVNPAATPADIYLOLSTFGT--LAAGSGTEIONRVKSDMSNFTETN 121
DB 737 IGASVNTFNFVKLSGVSADYLLVEGFSSGAGQLPGKDTDIDYVRFKNDMSVYNQAD 796
OY 122 DYSY-GTNTTFQDMTKRYTVYVGVLVWGTE 150

Db 797 DMSHLOSMTNYGENAKVTLTYDGYLVNGQE 826

Search completed: August 29, 2002, 16:37:47
Job time: 957 sec

This Page Blank (uspto)

Tue Sep 3 13:42:03 2002

us-09-917-384-5.ra1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:22:34 ; Search time 47.58 Seconds
(without alignments)
77.004 Million cell updates/sec

Title: US-09-917-384-5
Perfect score: 806
Sequence: 1 VSGGLKQYKKNNDSPGDNO.....FQDKTKTYVYVNGVLWGTE 150

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	367	45.5	551	2 US-09-033-537A-1	Sequence 1, Appl1
2	361.5	44.9	700	2 US-07-862-588B-2	Sequence 2, Appl1
3	359.5	44.6	493	4 US-09-198-956-10	Sequence 10, Appl1
4	359.5	44.6	493	4 US-09-198-955A-12	Sequence 12, Appl1
5	358.5	44.5	167	5 PCT-US95-13813-9	Sequence 9, Appl1
6	340	42.2	616	4 US-09-136-574A-47	Sequence 47, Appl1
7	339	42.1	1751	4 US-09-136-574A-44	Sequence 44, Appl1
8	336	41.7	1426	4 US-09-136-574A-43	Sequence 43, Appl1
9	329	28.4	162	1 US-08-048-164A-2	Sequence 2, Appl1
10	329	28.4	162	1 US-08-460-452-2	Sequence 2, Appl1
11	329	28.4	162	1 US-08-460-457-2	Sequence 2, Appl1
12	329	28.4	162	1 US-08-460-455-2	Sequence 2, Appl1
13	329	28.4	162	2 US-08-330-394A-2	Sequence 2, Appl1
14	329	28.4	163	3 US-09-006-636-7	Sequence 7, Appl1
15	329	28.4	163	3 US-09-006-632-7	Sequence 7, Appl1
16	329	28.4	163	4 US-09-277-716-22	Sequence 22, Appl1
17	329	28.4	163	4 US-09-330-394A-29	Sequence 29, Appl1
18	326	28.0	156	2 US-08-330-394A-22	Sequence 22, Appl1
19	326	28.0	531	2 US-07-862-588B-7	Sequence 7, Appl1
20	326	28.0	531	2 US-07-862-588B-7	Sequence 7, Appl1
21	81	10.0	238	3 US-09-286-690-7	Sequence 7, Appl1
22	81	10.0	238	3 US-09-286-691-23	Sequence 23, Appl1
23	80	9.9	326	4 US-09-687-147-23	Sequence 23, Appl1
24	79	9.8	2123	4 US-08-968-685A-10	Sequence 10, Appl1
25	78	9.7	221	1 US-08-315-695-16	Sequence 16, Appl1
26	78	9.7	237	1 US-08-103-998-4	Sequence 4, Appl1
27	78	9.7	237	1 US-08-103-998-4	Sequence 4, Appl1

28	78	9.7	434	2 US-09-008-962-1	Sequence 1, Appl1
29	78	9.7	434	2 US-08-675-507-1	Sequence 1, Appl1
30	78	9.7	434	3 US-09-213-205-1	Sequence 48, Appl1
31	78	9.7	2048	4 US-09-268-347-48	Sequence 5, Appl1
32	77.5	9.6	892	4 US-09-336-447A-5	Sequence 6, Appl1
33	77.5	9.6	140	1 US-08-476-275-6	Sequence 9, Appl1
34	77	9.6	1338	4 US-08-728-470-9	Sequence 10, Appl1
35	77	9.6	1338	4 US-08-719-641-9	Sequence 10, Appl1
36	77	9.6	1529	4 US-08-728-470-10	Sequence 9, Appl1
37	77	9.6	1529	4 US-08-719-641-10	Sequence 9, Appl1
38	77	9.6	1599	2 US-08-617-697-9	Sequence 10, Appl1
39	77	9.6	1600	2 US-08-617-697-10	Sequence 10, Appl1
40	76.5	9.5	211	1 US-07-842-349-2	Sequence 2, Appl1
41	76.5	9.5	211	1 US-08-246-686-2	Sequence 2, Appl1
42	76	9.4	625	2 US-08-532-547-7	Sequence 7, Appl1
43	76	9.4	625	2 US-08-532-547-9	Sequence 9, Appl1
44	76	9.4	625	2 US-09-019-809-7	Sequence 7, Appl1
45	76	9.4	625	4 US-09-019-809-9	Sequence 9, Appl1

ALIGNMENTS

RESULT 1
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sh lein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,537A
; FILING DATE: 02-MAR-1998
; CLASSIFICATION: 008
; PRIOR APPLICATION DATA: 0993/95
; APPLICATION NUMBER: 08-SEP-1995
; FILING DATE: 08-SEP-1995
; APPLICATION NUMBER: PCT/DK96/00364
; FILING DATE: 03-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4492,204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-033-537A-1

Query Match 45.5%; Score 367; DB 2; Length 551;
Best Local Similarity 47.0%; Pred. No. 2,1e-27;
Matches 71; Conservative 31; Mismatches 43; Indels 6; Gaps 4;

QY 2 SGGKLVQYKNNDSAPGDQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLYVNC 61
DB 402 TGNLVQYKVGSTASATDQMRSPFNKNGTTPVNLGSKLRYTTKD-GTDMASAFDM 460
QY 62 AAMCGNIRASFGSVNPATPTADTYLQLSFT--GGTLAAGSTGEIQNRKSDMSNTE 119
DB 461 A0IGASNSVAFF--ANFTGSNDTYVELSFGSAGSIPAGGTGDIQLMRKTDMSNENE 518
QY 120 TNDYGY-GTNTFQDMTKVTVYVNGVLWGT 149
DB 519 ANDISYDGAKTAYADMNRVTLHQNGTLWGT 549

RESULT 2
US-07-862-588B-2
; Sequence 2, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch Ieln, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T./ Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425, 204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-862-588B-2

Query Match 44.9%; Score 361.5; DB 2; Length 700;
Best Local Similarity 49.4%; Pred. No. 9,6e-27;
Matches 76; Conservative 29; Mismatches 42; Indels 7; Gaps 5;

QY 1 VSGGLKVOYKND-SAPGDQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLYVNC 59
DB 549 VNSDLVQYKDDRNATDQIKPFPHNQNGKTSPIVDLSLRLRYFTTD-SSAANNGMI 607

QY 60 DWAMCGGIRASFGSVNPATPTADTYLQLSFT--GCTLAAGSTGEIQNRKSDMSNF 117
DB 608 DWAKLGSSNIQISFGNHNA--DSDTYAELFGSSGAGSIEGGGEGIQLRMSKADMSNF 665
QY 118 TETNDYSY-GTNTFQDMTKVTVYVNGVLWGT 150
DB 666 NEANDYSFGAKTAYIADMNRVTLHQNGTLWGT 699

RESULT 3
US-09-198-956-10
; Sequence 10, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulten, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377, 200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
; US-09-198-956-10

Query Match 44.6%; Score 359.5; DB 4; Length 493;
Best Local Similarity 44.0%; Pred. No. 9,5e-27;
Matches 70; Conservative 34; Mismatches 44; Indels 11; Gaps 3;

QY 1 VSGGLKVOYKNNDSAPGDQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLYVNC 60
DB 331 VSGNLKVEFYNSPDDTSTINPQKVTNTGSSAIDSLTLRYVYVYDQKQTFW-CD 389
QY 61 WAM-----GCGNIRASFGSVNPATPTADTYLQLSFTGCTLAAGSTGEIQNRV 111
DB 390 HAAIGNSNGSYGITSNVGCTFVKMSSSTNNADTYLEISFTGTLPECAHV-QIOGRFAK 448
QY 112 SDMSNTEFTNDYSYGTNTFQDMTKVTVYVNGVLWGT 150
DB 449 NDMSNVTGSDYSFQRSQFVEMDQVATVANGVLWGRE 487

RESULT 4
US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulten, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schmitt, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580e1 Pectate Lyases
; FILE REFERENCE: 5378, 200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97

```

PRIORITY FILING DATE: 1997-11-24
PRIORITY APPLICATION NUMBER: 1344/97
PRIORITY FILING DATE: 1997-11-24
PRIORITY APPLICATION NUMBER: 60/067,249
PRIORITY FILING DATE: 1997-12-02
PRIORITY APPLICATION NUMBER: 60/067,240
PRIORITY FILING DATE: 1997-12-02
PRIORITY APPLICATION NUMBER: 09/073,684
PRIORITY FILING DATE: 1998-05-06
PRIORITY APPLICATION NUMBER: 09/184,217
PRIORITY FILING DATE: 1998-11-02
NUMBER OF SEQ ID NOS: 32
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 12
LENGTH: 493
TYPE: PRF
ORGANISM: Clostridium thermocellum
IS-09-198-955A-12

```

	Query Match	44.6%;	Score 359.5;	DB 4;	Length 493;	
	Best Local Similarity	44.0%;	Pred. No. 9.5e-27;		Indels 11;	Gaps 3;
	Matches	Conservative	34;	Mismatches 44;		
OY	1 VSGGLKVOYKKNDSAPGDNQIKPGLQLVNTGSSNDLSGVYRVWFTRDGGSSLLTYNCD	60				
	:: :::: : : :: :: :: :	389				
Db	331 VSGNKLEVFYSNSDDTTNSINPOFKTNGSSDAIDSLKLIRYYTVDGOKDOTFW-CD	111				
OY	61 WAAW-----GCCGTIRASFSGSVNPATPTADTLYLOLSTFGCTLAAGSITGEIONRYNK	448				
	::: :::: : :::: : :: :: :: :					
Db	390 HAAIIGSNCSNGITSNNKGTFEYSSNSTNADYILEISFTGGLTPGAHV-QIOGRRAK	150				
OY	112 SDMSNFTEINDSYGNTTFEDPMIKVIYVNGVLWGTE	487				
	: ::: :: :::: : :::: : :: :: :: :					
Db	449 NDMGNYQTOSMDYSFKRISOFEVMDQVTALINGVLWYKE					

RESULT 5
 PCT-US95/13813-9
 Sequence 9, Application PC/TUS9513813
 GENERAL INFORMATION:
 APPLICANT: Yeda Research and Development Co. Ltd.
 APPLICANT: Ramot University Authority for Applied
 Research and Industrial Development Ltd.
 APPLICANT: Technion Research and Development Foundation Ltd.
 APPLICANT: Bayer, Edward A.
 APPLICANT: Morey, Ely
 APPLICANT: Wilchek, Melf
 APPLICANT: Lamed, Raphael
 APPLICANT: Shoham, Yuval
 TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)
 TITLE OF INVENTION: PROTEINS AND USE THEREOF
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Brody and Nelmark
 STREET: 419 Seventh Street N.W., Ste. 300
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentln Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/13813
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Brody, Roger L.
 REGISTRATION NUMBER: 25,618
 REFERENCE/DOCKET NUMBER: BAYER-3

```

1      TELEX: 248633
2      INFORMATION FOR SEQ ID NO: 9:
3      SEQUENCE CHARACTERISTICS:
4      LENGTH: 167 amino acids
5      TYPE: amino acid
6      STRANDEDNESS: single
7      TOPOLOGY: linear
8      MOLECULE TYPE: peptide
9      PCT-US95-13813-9
10
11      length 167:

```

[illegible]

RESULT 6
 US-09-136-574A-47
 Sequence 47, Application US/09136574A
 Patent No. 6294366
 GENERAL INFORMATION:
 APPLICANT: Farrington, Graham K.
 Anderson, Paige
 Gibbs, Moreland
 Bergquist, Peter
 Daniels, Roy
 Morgan, Hugh W.
 Williams, Diane P.
 TITLE OF INVENTION: Compositions and Methods for
 Treating Cellulose Containing
 Cellulase Enzyme Compositions
 NUMBER OF SEQUENCES: 49
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Howson and Howson
 STREET: Spring House Corporate Center, P.O. Box 457
 CITY: Spring House
 STATE: PA
 COUNTRY: USA
 ZIP: 19477
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/136,574A
 FILING DATE: 19-Aug-1998
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION NUMBER: US 08/932,571
 APPLICATION DATE: September 19, 1997
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Bak, Mary E.
 REGISTRATION NUMBER: 31,215
 REFERENCE/DOCKET NUMBER: 1997US001/CIP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-540-9200
 TELEFAX: 215-540-5818
 TELEX: <Unknown>

;; INFORMATION FOR SEQ ID NO: 47:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 616 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-09-136-574A-47

Query Match 42.2%; Score 340; DB 4; Length 616;
Best Local Similarity 43.8%; Pred. No. 9.5e-25;
Matches 67; Conservative 30; Mismatches 52; Indels 4; Gaps 3;

OY 1 VSGIKVQTKNDAPGDNQIKPGIQLVNTGSSVDLSTVYRVYFTRDGGSSLTLYNCND 60
DB 1 MSGGKVVLYKNNETASGTSIRPWFKIVNGSSVDLSRKRIRYWTYDGDKPSAV-CD 59
OY 61 WAMGCGNIRASFGSVNPAFTADRYLQLSFTGCT--LAAGSGTEIGIONRVKSPMSNFT 118
DB 60 WAIQASNTFNFVKLSSGVSADYLYLEVGFSSGAGQLQPKDGTGDIQVRFNKNDMSNYN 119
OY 119 ETNDYSY-GTNTTFQDWTKVYVYNGVLWGTE 150
DB 120 QADWMSWLSMTNYGENAKVTILYVDGVLWGQE 152

RESULT 7
US-09-136-574A-44
; Sequence 44, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1751 amino acids

;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-136-574A-44

Query Match 42.1%; Score 339; DB 4; Length 1751;
Best Local Similarity 44.7%; Pred. No. 4.5e-24;
Matches 67; Conservative 29; Mismatches 50; Indels 4; Gaps 3;

OY 4 GLKVKYKNDASRGDNQIKPGIQLVNTGSSVDLSTVYRVYFTRDGGSSLTLYNCMDAA 63
DB 678 GKVKVLYKNNETASGTSIRPWFKIVNGSSVDLSRKRIRYWTYDGDKPSAV-CDWAQ 736
OY 64 MCGNIRASFGSVNPAFTADRYLQLSFTGCT--LAAGSGTEIGIONRVKSPMSNFTETN 121
DB 737 IGASNTFNFVKLSSGVSADYLYLEVGFSSGAGQLQPKDGTGDIQVRFNKNDMSNYNAD 796
OY 122 DYSY-GTNTTFQDWTKVYVYNGVLWGTE 150
DB 797 DWSWLSMTNYGENAKVTILYVDGVLWGQE 826

RESULT 8
US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1426 amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear

Tue Sep 3 13:42:03 2002

us-09-917-384-5.ral

MOLECULE TYPE: NO. 62943666
SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-09-136-574A-43

Query Match 41.7%: Score 336; DB 4; Length 1426;
Best Local Similarity 44.7%: Pred. No. 6,7e-24;
Matches 67; Conservative 28; Mismatches 51; Indels 4; Gaps 3;

QY 4 GLAVQYKNNDSAPGDNQIKPGQLVNTGSSVDLSTVYRVYRFTRDGSSSTLYVNCMDMA 63
DB 413 GLKVLKNNETSASTGSRPFKRYVGGSSVDLSRVKIRYIVDGRKQPSAY-CDMAQ 471
QY 64 MCGGNIRASGYNPATPTADTYLQSLFTGRT-LAAGSGTEIONRVNKSDFNFTETN 121
DB 472 IGASNVTFNFVKLSSGVADVLYLEVCFSSGAGQLQPKADGDIQVRNKNDSMYNQAD 531
QY 122 DYST-GTNTTFQDWTKYVYVNGVLYWGT 150
DB 532 DMSHLOSMTDYGNAKTYLYVDGVLWGOE 561

RESULT 9
US-08-048-164A-2
Sequence 2, Application US/08048164A

GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Shpiegl, Itai
APPLICANT: Goldstein, Marc A.
APPLICANT: Dol, Roy H.
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/048,164A
FILING DATE: 14-APR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-048-164A-2

Query Match 28.4%: Score 229; DB 1; Length 162;
Best Local Similarity 35.8%: Pred. No. 7.8e-15;
Matches 58; Conservative 31; Mismatches 55; Indels 18; Gaps 7;
QY 2 SGLKQYKNNDSAPGDNQIKPGQLVNTGSSVDLSTVYRVYRFTRDGSSSTLYVNCMD 61
DB 3 TSMKSYEFYNSKNSAQTNSITPIIKITNTSDSLNLDNVKRYIYITSDGRTGQTFW-CDH 61

QY 62 AAMCGN-----IRASF--GSYNPATPTADTYLQSLFTGRTLAAGSGTGE---IQNR 108
DB 62 AGALLGNSYVDNTSKYTANFVKETASP-TSTYDIVERGFASG--AAVLKKGQFTTIOGR 118
QY 109 VNKSDFNFTETNDYST-GTNTTFQDWTKYVYVNGVLYWGT 149
DB 119 ITKSDMSNTYTQNDYSPDASSTPVYVNPVYTGIGAKVLYGT 160

RESULT 10
US-08-460-462-2
Sequence 2, Application US/08460462

GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Shpiegl, Itai
APPLICANT: Goldstein, Marc A.
APPLICANT: Dol, Roy H.
TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,462
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-460-462-2

Query Match 28.4%: Score 229; DB 1; Length 162;
Best Local Similarity 35.8%: Pred. No. 7.8e-15;
Matches 58; Conservative 31; Mismatches 55; Indels 18; Gaps 7;
QY 2 SGLKQYKNNDSAPGDNQIKPGQLVNTGSSVDLSTVYRVYRFTRDGSSSTLYVNCMD 61
DB 3 TSMKSYEFYNSKNSAQTNSITPIIKITNTSDSLNLDNVKRYIYITSDGRTGQTFW-CDH 61
QY 62 AAMCGN-----IRASF--GSYNPATPTADTYLQSLFTGRTLAAGSGTGE---IQNR 108
DB 62 AGALLGNSYVDNTSKYTANFVKETASP-TSTYDIVERGFASG--AAVLKKGQFTTIOGR 118
QY 109 VNKSDFNFTETNDYST-GTNTTFQDWTKYVYVNGVLYWGT 149
DB 119 ITKSDMSNTYTQNDYSPDASSTPVYVNPVYTGIGAKVLYGT 160

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,455
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 869-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-460-455-2

Query Match 28.4%; Score 229; DB 2; Length 162;
Best Local Similarity 35.8%; Pred. No. 7,8e-15;

Matches 58; Conservative 31; Mismatches 55; Indels 18; Gaps 7;

QY 2 SGGKAYKNNDSAPGDNQIKPGQLVNTGSSVDLSTVYRYVFTFRDGGSTLYVNCW 61
DB 3 TSSMSVEFYNSKNSAQTNSTPIIKITNTSDSLNLDVRYVYTSQGGQTFW-CDH 61
QY 62 AAMGCN-----IRASF--GSVNPATPTADTYLQLSFGTLAGSGTGE---IQNR 108
DB 62 AGALGNSYVDNISKYANVKETASP-TSTYDYVEFGFASG--AATLKKGQFTIQGR 118
QY 109 VNKSDFSNETNDYSY-GTNTTFQDMTKYVYVNGVLVWGT 149
DB 119 ITRSDMSNTQTNDYSFSDASSSTPVVNPVKYTGIGAKVLT 160

RESULT 14

US-08-330-394A-2

Sequence 2, Application US/08330394A

Patent No. 5856201

GENERAL INFORMATION:

APPLICANT: Shoseyov, Oded

APPLICANT: Yosef, Karney

APPLICANT: Shpiegel, Itai

APPLICANT: Goldstein, Marc A.

APPLICANT: Dol, Roy H.

TITLE OF INVENTION: METHODS OF DETECTION USING THE

TITLE OF INVENTION: CELLULOSE BINDING DOMAIN

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 AVENUE OF THE AMERICAS

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/330,394A

FILING DATE: 27-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CIP OF PCT/US94/04132
FILING DATE: 14-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: MISTROCK, S. LESLIE
REFERENCE/DOCKET NUMBER: 7809-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66441 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-330-394A-2

Query Match 28.4%; Score 229; DB 2; Length 162;

Best Local Similarity 35.8%; Pred. No. 7,8e-15;

Matches 58; Conservative 31; Mismatches 55; Indels 18; Gaps 7;

QY 2 SGGKAYKNNDSAPGDNQIKPGQLVNTGSSVDLSTVYRYVFTFRDGGSTLYVNCW 61
DB 3 TSSMSVEFYNSKNSAQTNSTPIIKITNTSDSLNLDVRYVYTSQGGQTFW-CDH 61
QY 62 AAMGCN-----IRASF--GSVNPATPTADTYLQLSFGTLAGSGTGE---IQNR 108
DB 62 AGALGNSYVDNISKYANVKETASP-TSTYDYVEFGFASG--AATLKKGQFTIQGR 118
QY 109 VNKSDFSNETNDYSY-GTNTTFQDMTKYVYVNGVLVWGT 149
DB 119 ITRSDMSNTQTNDYSFSDASSSTPVVNPVKYTGIGAKVLT 160

RESULT 15

US-09-006-636-7

Sequence 7, Application US/09006636

Patent No. 6005092

GENERAL INFORMATION:

APPLICANT: Shoseyov, Oded

APPLICANT: Shani, Ziv

TITLE OF INVENTION: ARABIDOPSIS THALIANA ENDO-1,4-BETA-

TITLE OF INVENTION: GLUCANASE GENE, PROMOTER AND PROTEIN

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/006,636

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Baldwin, Geraldine F.

REGISTRATION NUMBER: 31,232

REFERENCE/DOCKET NUMBER: 7809-019

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 163 amino acids

TYPE: amino acid

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:21:43 : Search time 127 Seconds
(without alignments)
912.204 Million cell updates/sec

Title: US-09-917-384-6
Sequence: 5536
1 MERTQOSGRNCRVGRGTRM.....GFNGSYSGTNAPATLSCTAS 1043

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1023	18.5	1010	AAW34989	Teredinibacter end
2	1000	18.1	596	AAW34989	Thermotoga cellu
3	999	18.0	423	AAW34989	Thermotoga cellu
4	646.5	11.7	381	AAW06368	Streptomyces livid
5	646.5	11.7	381	AAW06368	Rhodothermus marin
6	637	11.5	386	AAW67497	prolin sequence o
7	635.5	11.5	371	AAW06367	Streptomyces sp. E
8	635.5	11.5	371	AAW06367	Streptomyces livid
9	635.5	11.5	371	AAW06367	Amino acid sequenc
10	635.5	11.5	429	AAW84346	Amino acid sequenc
11	633.5	11.4	371	AAW08473	Actinomyces sp. 3

12	633.5	11.4	371	AAW67496	Actinomyces cell
13	574.5	10.4	395	AAW84797	Amino acid sequenc
14	573	10.4	473	AAW01077	H. insolens cellu
15	565.5	10.2	476	AAW44853	Humicola insolens
16	565.5	10.2	476	AAW44853	H. insolens cellu
17	548	9.9	471	AAW50308	Cellulohydrolyase
18	540	9.8	471	AAW77262	T. longibrachiatum
19	540	9.8	471	AAW02025	Trichoderma cellu
20	539	9.7	457	AAW25789	Acetomoniun cellu
21	472.5	8.5	446	AAW01076	H. insolens cellu
22	467	8.4	1751	AAW13493	Truncated cellulas
23	449	8.1	1426	AAW13492	Truncated cellulas
24	444.5	8.0	449	AAW56739	Orpinomyces cellu
25	431.5	7.8	432	AAW18414	Amino acid sequenc
26	417.5	7.5	551	AAW18790	Corrected Bacillus
27	414	7.5	562	AAW69508	Acidothermus cellu
28	413	7.5	1853	AAW43108	C. thermocellum ce
29	412	7.4	459	AAW6738	Orpinomyces cellu
30	410.5	7.4	521	AAW8788	Acidothermus marin
31	407.5	7.4	260	AAW06369	Rhodothermus deserto
32	407.5	7.4	260	AAW14882	Emeritella deserto
33	407.5	7.4	260	AAW14882	Amino acid sequenc
34	406	7.3	521	AAW8927	A. cellulolyticus
35	405	7.3	782	AAW15625	Acidothermus cellu
36	404.5	7.3	521	AAW48787	Cellulase AE-1. A
37	404.5	7.1	476	AAW54123	Acidothermus cellu
38	393	7.0	700	AAW13227	Novel endoglucanas
39	388.5	6.9	499	AAW42122	Novel endoglucanas
40	383	6.9	531	AAW15238	Scaffoldin protein
41	381.5	6.9	1352	AAW63962	Amino acid sequenc
42	377	6.8	616	AAW13494	Truncated cellulas
43	376	6.7	1350	AAW63963	Amino acid sequenc
44	373.5	6.7	493	AAW28850	Pectate lyase-link
45	371.5	6.7	20	AAW28850	

ALIGNMENTS

RESULT 1
AAW34989 standard: Protein; 1010 AA.
ID AAW34989
AC AAW34989:
XX
XX 21-MAY-1998 (first entry)
DE Teredinibacter endoglucanase.
XX
XX Endoglucanase: cellulase; carboxymethylcellulose; cellulose;
XX Endoglucanase: beta-1,4-glycosidic bond; hydrolysis; saccharification;
XX biomass; thermophilic; glycosidase.
XX
XX Teredinibacter sp. (Clone 426P1).
OS
XX
XX W09744361-A1.
PN
XX 27-NOV-1997.
PD
XX
XX 22-MAY-1997; 97WO-US08793.
PE
XX
XX 22-MAY-1996; 96US-0651572.
PR
XX
XX (RECO-) RECOMBINANT BIOCATALYSIS INC.
PA
XX Lam DE, Mathur EJ;
PI
XX WPI: 1998-018435/02.
DR N-PSDB: AAT94197.
XX
XX Endoglucanase(s), preferably form archaeal bacterium, AEP1 1a
XX useful to degrade carboxymethylcellulose and hydrolyse of
XX beta-1,4-glycosidic bonds in cellulose
PT

```

XX PS Claim 1; Fig 1E; 164pp; English.
XX CC This protein comprises an endoglucanase of Terebrinibacter (clone
CC 432P1) that is capable of degrading carboxymethylcellulose and of
CC hydrolysing the beta-1,4-glycosidic bonds in cellulose. It has
CC homology to an endoglucanase of archaeobacterium APB11a (see
CC AAM34985). It can be produced from native cells or from recombinant
CC host cells, especially prokaryotic host cells transformed with a
CC plasmid or virus-derived vector including the endoglucanase DNA
CC (see AAT94197). 24 Endoglucanases (see AAM34986-W35008) are claimed.
CC They can be used to degrade cellulose for the conversion of plant
CC biomass into fuels and chemicals, for use in detergents, textiles,
CC animal feed, waste treatment, and in the fruit juice and brewing
CC industries for the clarification and extraction of juices.
XX SQ Sequence 1010 AA;

Query Match 18.5%; Score 1023; DB 19; Length 1010;
Best Local Similarity 48.5%; Pred. No. 7.3e-57;
Matches 210; Conservative 68; Mismatches 121; Indels 34; Gaps 13;

OY 57 VDNPPAGA-TFEVNPYMAOEVOSEANOTNATLAKMRVSYSTAVWMDRTAINGVNG 115
DB 598 vdnppagaqkylinpms-----asaaneppgs-----vianeppfwmndrigalepad 647
OY 116 GPGLTYYIDALISQOOGTTPPEVIEIYIDLPGRCALALASNGELPATAAGLQTYFTQYID 175
DB 648 gmgldrhinealaq----gadlfmfvvydlpurdcaalangelrlsedgfnlysdyla 703
OY 176 PLASTLSPKYSLSLIVYIIIEPDSLPAVNTNMSIQAC--ATNVPYEEGIEVALTKLNI 233
DB 704 plvelslpypaygiklaavleavslpnlvtlnsepdceangpgyrdgtrlnatelnlqk 763
OY 234 PNVTIYMDAAHSGWLGPNN-ASGYVOEYOKYLNASIGVNGIDGVTNANTPTKEPPM 292
DB 764 pnvtyvdalshgvlgnwdfnfgynlllyevanlgsnglnpnlagtrnsanlytpreepfl 823
OY 233 -TATQOVGGQPYESANFYQWNPDIADYAVDLVSLVAAGFPSSIGMLITLNGWGP 351
DB 824 pdanlqvgqpyrsdlyewmsylaekepfvltwrsamlskmpsslgnlidlarnwgpp 883
OY 352 NEPTGPSTATDVNTFVNOSKIDLRHGLCMONGAGLGOPPOASPTDEPNHLDAYWI 411
DB 884 erptagsismnltfvesrlidrtelnrgwcnqpg-ygyytrpta--pspgldayvvw 938
OY 412 KPGESEDGTSNAA--DPT-TGKKSDEPMCDP---TYTTSYGLTNALPNSIAGOMFPA 463
DB 939 kpggesdgvspnfelidpnudpkqhdpmcdpfaansnsayg--tgampnaphagrwlpe 996
OY 464 QPDOLVANARPAV 476
DB 997 aqqlllenayppl 1009

RESULT 2
AAR90715
ID AAR90715 standard; Protein, 596 AA.
XX AC AAR90715;
XX DT 06-APR-1996 (first entry)
XX DE Thermostable cellulase-E3.
XX KW Cellulase-E3; Thermomonospora fusca; thermostable; plasmid pS24;
KW plasmid pS26; Escherichia coli; cloning; Streptomyces lividans;
KW papain; catalytic domain; cellulose; chitosan; process resistance;
KW synergism; cellobiohydrolase; beta-glucosidase; saccharification;
KW surfactant additive; paper recycling; delinking; paper refining.
OS Thermomonospora fusca strain YX36.

```

```

XX FH Key
XX FT Peptide
XX FT Peptide
XX FT Peptide
XX FT Domain
XX PN W09600281-A1.
XX 04-JAN-1996.
XX 23-JUN-1995; 95WO-US09069.
XX 24-JUN-1994; 94US-0265429.
XX (CORR ) CORNELL RES FOUND INC.
XX Walker LP, Wilson DB, Zhang S;
XX WPI, 1996-068865/07.
XX N-PSDB; AAT15595.
XX Cellulase from Thermomonospora fusca - is thermostable and is useful
XX in a variety of industrial applications e.g. clarification of fruit
XX juices, fabric softening, etc
XX Claim 1; Page 32-37; 53pp; English.
XX The sequence represents thermostable cellulase-E3 (EC-3.2.1.4) from
XX Thermomonospora fusca, and is encoded by inserts in plasmid pS24
XX and plasmid pS26, obtained by cloning in Escherichia coli. The
XX gene may be cloned and expressed in E. coli or Streptomyces
XX lividans for recombinant cellulase-E3 production. Cellulase-E3,
XX or a catalytically active polypeptide obtained by papain digestion,
XX hydrolyses cellulose or chitosan at pH 5-11 and 40-70 deg C, and
XX has significant activity at over 60 deg C. E3 has higher stability
XX to proteolysis in culture supernatant than T. fusca cellulases-E2
XX and -E5, and shows strong synergistic activity when combined with
XX other cellulases, cellobiohydrolase and optionally beta-glucosidase.
XX The cellulase and mixtures may be used e.g. in cellulose
XX saccharification for ethanol production, fruit juice clarification,
XX as a surfactant additive, or in delinking or refining of recycled
XX paper.
XX SQ Sequence 596 AA;

Query Match 18.1%; Score 1000; DB 17; Length 596;
Best Local Similarity 47.7%; Pred. No. 1.1e-55;
Matches 210; Conservative 55; Mismatches 143; Indels 32; Gaps 9;

OY 50 PAIAATHVNDPPAGATFEVNPYMAOEVOSEANOTNATLAKMRVSYSTAVWMDRTA 109
DB 172 punpekevpndieagklvynpwsakaapepgs-----avaestavldiriga 221
OY 110 ING---VNGGPGLTYYIDALISQOOGTTPPEVIEIYIDLPGRCALALASNGELPATAAG 165
DB 222 legndsapltgsmglrdhleeav-rqsggdplrtlgvlynlpurdcaalangelrldpe-- 278
OY 166 LOTYETQYIDPIASTLSN-PRYSSLSRLIVYIIIEPDSLPAVNTNMSIQACATAVPY----- 218
DB 279 ldrkyseyldpladlmwdfadyenlrivallleidsjplnvtngvgngfclcaymkqng 338
OY 219 YEGGIEYALTKHAIPNVYIYMDAAHSGWLGPNNASGYVOEYOKYLNAS-IGVNGIDOF 277
DB 339 ymgvgyalrkigelpnyvnyidaahngwlnfsgpsvdlfyeaanaagstvyvngf 398
OY 278 VNTVNTNPTPLKEPFMTATQOVGGQPYESANFYQWNPDIADYAVDLVSLVAAGFPSSIG 337
DB 399 lantanyasatvepyldvngtvgqlirgskvwdwngyvdeisfvqdlrqaallakfrsdl 458

```

OY	338	GMLIDTLRNGGNGNEPTGSPSTADVTMTFVNQSKIDLRHRRGLMCNONGGLGQPPQASP	397
DB	459	gmllidtlrnggngnrcrlpgssscldrlvdeesldrrllhpgnawcgaqslgetrvvp	518
OY	398	TDFPNALHDAVWVKPGESEKCTSAASADPTTGKKSDMCDPTVWTSGVLTN--ALPNS	454
DB	519	a----pvydayvwvkprgesdgaeeelpndegkyfdmcdpctlygnarlnnpsgalpna	574
OY	455	PIAGOWFPAGFDQLVANAPR	474
DB	575	plsghwfsaqfrelleanayp	594
RESULT 3			
AAAR0716			
ID	AAAR0716	standard; Protein: 423 AA.	
XX			
XX	AAAR0716;		
DT	06-APR-1996	(first entry)	
XX			
DE	Thermostable cellulase-E3 catalytic domain.		
KW	Cellulase-E3; Thermomonospora fusca; thermostable; papain;		
KW	catalytic domain; Escherichia coli; cloning; Streptomyces lividans;		
KW	cellulose; chitosan; protease resistance; synergism;		
KW	cellobiohydrolase; beta-glucosidase; saccharification;		
KW	surfactant additive; paper recycling; deinking; paper refining.		
XX			
OS	Thermomonospora fusca strain YX36.		
XX			
PN	WO9600281-A1.		
PD	04-JAN-1996.		
XX			
PF	23-JUN-1995; 95WO-0509069.		
XX			
PR	24-JUN-1994; 94US-0265429.		
XX			
PA	(CORR) CORNELL RES FOUND INC.		
XX			
PI	Walker LP, Wilson DB, Zhang S;		
XX			
DR	WPI; 1996-068865/07.		
DR	N-PSDB; AAT15596.		
XX			
PT	Cellulase from Thermomonospora fusca - is thermostable and is useful		
PT	in a variety of industrial applications e.g. clarification of fruit		
PT	juices, fabric softening, etc		
XX			
PS	Claim 1; Page 38-40; 53pp; English.		
XX			
CC	The sequence corresponds to a catalytically active domain of		
CC	thermostable cellulase-E3 (EC-3.2.1.4) from Thermomonospora		
CC	fusca, and is isolated by papain cleavage of the full-length		
CC	protein (AAAR0715). The binding domain and linker region have		
CC	been removed from the full-length protein. The protein may be		
CC	produced in recombinant form in Escherichia coli or Streptomyces		
CC	lividans. The domain retains full activity, but shows reduced		
CC	cellulose binding. The catalytic domain, or the corresponding		
CC	full-length cellulase, hydrolyses cellulose or chitosan at pH		
CC	5-11 and 40-70 deg C, and has significant activity at over 60		
CC	deg C. E3 has higher stability to proteolysis in culture		
CC	supernatant than T. fusca cellulases-E2 and -E5, and shows strong		
CC	synergistic activity when combined with other cellulases,		
CC	cellobiohydrolase and optionally beta-glucosidase. The cellulase		
CC	and mixtures may be used e.g. in cellulose saccharification for		
CC	ethanol production, fruit juice clarification, as a surfactant		
CC	additive, or in deinking or refining of recycled paper.		
XX			
Sequence	423 AA;		

Query Match	18.0%	Score 999	DB 17	Length 423
Best Local Similarity	48.3%	Pred. No. 7.7e-56		
Matches 209	Conservative 55	Mismatches 137	Indels 32	Gaps 9
QY	57	VDNPRAGATFFPNPWAQVEQSEANQGNATLAAMRVSTYSTFVMMDRIAING----	112	
Db	6	vdnpragaklyynprwsakaadepggs-----avanesstavwldrlgalsgndsp	55	
QY	113	VNGSGELTYYUDDAALSOQOGTTPPEYIEIYIDLPGRDCAALASNEELPATAGLTQYETQ	172	
Db	56	ttsgmsgrldchleear-rsgsgdrlclqyynlgnlprdcalaasngelppde--ldrylse	112	
QY	173	YIDPIASILSN-PKYSLSRLVTIIEPDSLPANAVTKMSIOACATAVPY-----YEOGIEY	225	
Db	113	yldpradlmwfdiagyeirlvalleldslrlnltnvngsgfclcaymkqngyynvnygy	172	
QY	226	ALTKHAALPNYIYDDAAHSGWLGRPNNASGYQVEQKVLNMS-IGVNGIDGFYNTNTAY	284	
Db	173	alktgeiprvnyyidaahgwliyvdsanfgpsvdlfyeaanaagstvdynvngflsantany	232	
QY	285	TPLEKPEMTATQOVCGQVESANFQOMPPDIDEDAYUAYDLVSRLLAACGSSIGMLIPTL	344	
Db	223	satvepyldnngltvngqllrsgskwdmngyvdehsfvgdltgallakgsfrsldnlltdts	292	
QY	345	RNGWGNPEPTGPRSTATVNTFFVNOSKIDLRORHGLMCONCAGLGGORPOASPTDFPNAH	404	
Db	293	rngwsgprnrtpgspsstldntlyvdesrldtrllhpgnwcngcgauglgeprtpvna----pg	348	
QY	405	LDAYWIRPQGSDDGTSAAADPTTKKSDPMCKDPIYTTTSTYGVLLN---ALPNSPLAGQWF	461	
Db	349	vdaywvwrkprgsdgaaseelpndegkfgtdfcmcdprcygnatrgnnpagallnaprlsgwlf	408	
QY	462	PAQFDQILVANARP	474	
Db	409	sattflllanayp	421	

RESULT	4
AA06368	
ID	AA06368 standard; Protein; 381 AA.
XX	
AC	AA06368;
XX	
DT	06-SEP-1999 (first entry)
XX	
DE	Streptomyces lividans Egitr-like cellulase.
XX	
KW	Cellulase; endoglucanase; Egitr; textile; feed additive; baking;
KW	food processing; grain wet milling; pulp; paper.
XX	
OS	Streptomyces lividans.
XX	
PN	MO9931255-A2.
XX	
PD	24-JUN-1999.
XX	
PE	14-DEC-1998; 98WO-US26552.
XX	
PR	16-DEC-1997; 97US-0991720.
XX	
PA	(GEMV) GENENCOR INT INC.
XX	
PI	Bower BS, Fowler T, Phillips JI;
XX	
DR	WPI; 1999-395187/33.
XX	
PT	Egitr like cellulase
XX	
PS	Example; Fig 6; 47pp; English.
CC	
CC	The present polypeptide represents a full-length sequence of a
CC	novel Egitr-like cellulase of Streptomyces lividans (celB). It was
CC	deduced from a gene sequence isolated from genomic DNA using PCR

primers (see AAY59180-91) based on conserved motifs (see AAY06325-29)
CC of Trichoderma reesei Egit-like enzymes and related enzymes. PCR
CC has been used to identify novel Egit-like enzymes, including the
CC present protein, from bacterial and fungal sources (see AAY06331-70).
CC The sequence shows homology to T. reesei Egit (see AAY06330). Also
CC provided by the invention are vectors, host cells and methods
CC for the recombinant production of such enzymes, which can be used
CC in the treatment of cellulose-containing textiles, as feed of
CC additives, in the treatment of wood pulp, in the reduction of
CC biomass to glucose, in the stone washing of indigo dyed denim, or
CC as laundry detergent components (all claimed).
SQ Sequence 381 AA:

Query Match 11.7%; Score 646.5; DB 20; Length 381;
Best Local Similarity 40.7%; Pred. No. 2.4e-33;
Matches 155; Conservative 45; Mismatches 144; Indels 37; Gaps 14;

684 GLAAG-----TAYTYVAAVDAAGNTSAPSTPVDCTPGPNONGVMSVODGEYRVQTNEMN 738
13 gllaa1gavlaafalvsslvtaapadqtl--lcep---fgtltlq-gryvvgqnnrwg 65

739 SSAOQCLTINTATGAMTVSTANFSGTGAPATYPSIYKCHMGNCTTKNGMPIDISQI 798
66 stapqcvta-ldtg-fryvqadgsaplngapksypsvfngchylnc-spgtdlprvldtv 122

799 GSATVSMSTTOVSSGAYDVAIDWNTSPPTTGGPNGTEIMIMLSRGGVOPFGSGTARG 858
123 saapsislygfydgavynasylwldptarldg-vnqtelimlfnrvspglqigsrvgt- 180

859 VYVAGHTMVMWMOGQQTSMKIIISVLTLPGATISNLDKRIFDAARGLNSDYLIDVE 918
181 asvgrtlevewsgngsndvlsfv-apsalsgsvfmdfvra1varglaendwyltsvq 239

919 AGEFIWOGGQGLGNSFSVSFV---TSGGV-----ACRATVYVNSDMGSGFATVTV 966
240 adfepwqngaglavnsfsstveltpgtdpdpdpgspacavsyntv-wqdgftadv 298

967 TMTGSRATNGMTVAMSGNGOTVNTWNTALTQSGASVTATNLISNNV1QPGOSTTGFN 1026
299 tntglapvqgwqlatflpsqgrltanwnasltpsgsvvlatgashnarlapsglsfgr 358

1027 GSYSGTNAAPT-----LSC 1041
359 gcygga1faeptgfrlntact 379

DB 359 gcygga1faeptgfrlntact 379

RESULT 5
AAB14881
ID AAB14881 standard: Protein; 381 AA.
XX
AC AAB14881;
XX
DT 21-NOV-2000 (first entry)
XX
DE Rhodothermus marinus Egit-like cellulase.
XX
KW Rhodothermus marinus; Trichoderma reesei; endoglucanase III; EgitII;
KM cellulase; mutant; enzyme stability; textile treatment;
KV wood pulp treatment; feed additive; detergent.
XX
OS Rhodothermus marinus.
XX
PN WO200037614-A2.
XX
PD 29-JUN-2000.
XX
PF 12-NOV-1999; 99WO-US26704.
XX
PR 18-DEC-1998; 98US-0216295.
XX
PA (GENEV) GENENCOR INT INC.

XX
PI Mitchinson C, Wendt DJ;
XX
DR WPI, 2000-482483/42.
XX
PT Novel endoglucanase III or endoglucanase III-like cellulase useful for
PT treating textiles and wood pulp comprises a substitution or deletion at
PT specified positions in the wild form of endoglucanase III -
XX
XX
PS Example 1; Fig 3; 52pp; English.
XX
CC The present sequence is a cellulase related to endoglucanase III (EgitII)
CC from Trichoderma reesei. Egit-like genes were isolated from genomic DNA
CC libraries constructed from various microorganisms by PCR. The isolated
CC genes showed significant homology to EgitII from T. reesei. Certain
CC substitution and deletion mutations have been incorporated into EgitII and
CC EgitII-like cellulases to produce variant enzymes with improved stability,
CC e.g. increased resistance to temperature stress. The mutants may be used
CC in textile and wood pulp treatment, as a feed additive, and for reducing
CC biomass to glucose. They are also useful for stonewashing or indigo dyed
CC denim and as an agent in laundry and dish detergents.
SQ Sequence 381 AA:

Query Match 11.7%; Score 646.5; DB 21; Length 381;
Best Local Similarity 40.7%; Pred. No. 2.4e-33;
Matches 155; Conservative 45; Mismatches 144; Indels 37; Gaps 14;

684 GLAAG-----TAYTYVAAVDAAGNTSAPSTPVDCTPGPNONGVMSVODGEYRVQTNEMN 738
13 gllaa1gavlaafalvsslvtaapadqtl--lcep---fgtltlq-gryvvgqnnrwg 65

739 SSAOQCLTINTATGAMTVSTANFSGTGAPATYPSIYKCHMGNCTTKNGMPIDISQI 798
66 stapqcvta-ldtg-fryvqadgsaplngapksypsvfngchylnc-spgtdlprvldtv 122

799 GSATVSMSTTOVSSGAYDVAIDWNTSPPTTGGPNGTEIMIMLSRGGVOPFGSGTARG 858
123 saapsislygfydgavynasylwldptarldg-vnqtelimlfnrvspglqigsrvgt- 180

859 VYVAGHTMVMWMOGQQTSMKIIISVLTLPGATISNLDKRIFDAARGLNSDYLIDVE 918
181 asvgrtlevewsgngsndvlsfv-apsalsgsvfmdfvra1varglaendwyltsvq 239

919 AGEFIWOGGQGLGNSFSVSFV---TSGGV-----ACRATVYVNSDMGSGFATVTV 966
240 adfepwqngaglavnsfsstveltpgtdpdpdpgspacavsyntv-wqdgftadv 298

967 TMTGSRATNGMTVAMSGNGOTVNTWNTALTQSGASVTATNLISNNV1QPGOSTTGFN 1026
299 tntglapvqgwqlatflpsqgrltanwnasltpsgsvvlatgashnarlapsglsfgr 358

1027 GSYSGTNAAPT-----LSC 1041
359 gcygga1faeptgfrlntact 379

DB 359 gcygga1faeptgfrlntact 379

RESULT 6
AAY67497
ID AAY67497 standard: Protein; 386 AA.
XX
AC AAY67497;
XX
DT 19-MAY-2000 (first entry)
XX
DE Protein sequence of Cella and cellulase 11AG8 fusion.
XX
KW Cellulase; Actinomycetes; detergent; feed additive; textile treatment;
KM pulp; paper; cellulase 11AG8.
XX
OS Streptomyces lividans.
XX
PA Synthetic.

XX Key Location/Qualifiers
 PH Peptide 1..46
 FT /note= "celA signal sequence"
 FT 47..386
 FT Protein /note= "cellulase 11A68 mature peptide"
 FT
 PN WO200009707-A1.
 XX
 PD 24-FEB-2000.
 XX
 PF 28-MAY-1999; 99WO-US11971.
 XX
 PR 24-JUN-1998; 98US-0104308.
 PR 18-NOV-1998; 98MO-US24649.
 PR 28-MAY-1999; 99US-0321981.
 XX
 PA (GEMV) GENENCOR INT INC.
 PI Jones BE, Van Der Kleij WAH, Van Solingen P, Weyler W;
 DR WPI: 2000-224344/19.
 DR N-PSDB: AAZ57031.
 XX
 PT A novel Actinomycetes cellulase and related DNA, useful for detergent
 PT compositions, treating textiles and paper or pulp
 XX
 PS Example 6; Fig 15; 72pp; English.
 CC The invention provides a cellulase from Actinomycetes. The cellulase can
 CC be used in a detergent composition, as an additive for animal feed and
 CC for the treatment of textiles or pulp and paper. The DNA encoding the
 CC cellulase can be used to identify homologous cellulases and for
 CC recombinant production of cellulases. The present sequence represents
 CC the protein sequence of a celA signal sequence and cellulase 11A68
 CC the fusion sequence contained in the expression cassette consisting of the
 CC GI promoter, celA signal sequence, cellulase 11A68 and GI terminator
 CC
 XX Sequence 386 AA:
 SQ
 Query Match 11.5%; Score 637; DB 21; Length 386;
 Best Local Similarity 40.0%; Pred. No. 9.8e-33;
 Matches 149; Conservative 48; Mismatches 134; Indels 36; Gaps 15;
 QY 698 VDAAGNTSAPSTPYDCTPPGNQ-----GVTSDGGEYRQVTENMSSAQCCLTINTANG 752
 DB 31 vs1vg1talagpaga-----nqgldrygtltlqg-ryvvgnmrwgtatqgc--lnvgn 83
 QY 753 AMVTSTANFSGGTGAPATYPSITKGCWGNCTTKNVGMPIQISQISAVTSMTTOVSS 812
 DB 84 gfe1qadqsvplngapksydvqchyncaprtt-1pmrls1s1gsapsvsvrytgn 142
 QY 813 GADVAYDIDWNTSTPTTGGOPNGTEIMTWLNSRGVOPFGSQRTGTVAGHTWNNWQO 872
 DB 143 gvyuaeyd1wldptptng-vnrteim1wfnrvpvgp1gsyvgt-ahvgsrsewvrgs 200
 QY 873 QTSWKIISYVLTFGATISINLDKATFADAAAGSLNTSD-YLDDVEAGFEIWOGGGIG 931
 DB 201 ngandv1sf-1apsalswsvfvdv-fvdqavshglapdwyl1ta1sgfepwge9tga 258
 QY 932 SNSFSVSVTSQG-----VACRAITYVNSDMGSGFATYATVINTGSRAVNGTWVMS 982
 DB 259 vns1saavnaagngt1p1sp1paqcyv1s1ht-1wp1g1fvd1t1cln1t1gs1p1d1f1t 317
 QY 983 FCGMOTYTNWNTALNQSASVYATNLYSN-NV1QPGQSTTFGPNCSYSGTN-AAPT--- 1037
 DB 318 1p1ag1t1v1sa1w1a11sp1sg1av1ar1st1g1s1a1ng1t1g1sf1g1t1s1g1f1n1p1a1g1r 377
 QY 1038 ---LSCT 1041
 DB 378 lngtact 384

RESULT 7
 ID AAY06367 standard; Protein; 371 AA.
 AC AAY06367;
 DT 06-SEP-1999 (first entry)
 DE Streptomyces sp. EgIII-like cellulase.
 KW Cellulase; endoglucanase; EgIII; textile; feed additive; baking;
 KW food processing; grain wet milling; pulp; paper.
 OS Streptomyces sp.
 PN WO931255-A2.
 PD 24-JUN-1999.
 PF 14-DEC-1998; 98MO-US26552.
 PR 16-DEC-1997; 97US-0991720.
 PA (GEMV) GENENCOR INT INC.
 PI Bower BS, Fowler T, Phillips JT;
 DR WPI: 1999-395187/33.
 XX
 PT EgIII like cellulase
 XX
 PS Example: Fig 6; 47pp; English.
 CC The present polypeptide represents a full-length sequence of a
 CC novel EgIII-like cellulase of Streptomyces sp. 11A68. It was
 CC deduced from a gene sequence isolated from genomic DNA using PCR
 CC primers (see AAX59180-91) based on conserved motifs (see AAY06325-29)
 CC of Trichoderma reesei EgIII cellulase and related enzymes. PCR
 CC has been used to identify novel EgIII-like enzymes, including the
 CC present protein, from bacterial and fungal sources (see AAY06331-70).
 CC The sequence shows homology to T. reesei EgIII (see AAY06330). Also
 CC provided by the invention are vectors, host cells and methods
 CC for the recombinant production of such enzymes, which can be used
 CC in the treatment of cellulose-containing textiles, as feed
 CC additives, in the treatment of wood pulp, in the reduction of
 CC biomass to glucose, in the stone washing of indigo dyed denim, or
 CC as laundry detergent components (all claimed).
 CC
 XX Sequence 371 AA:
 SQ
 Query Match 11.5%; Score 635.5; DB 20; Length 371;
 Best Local Similarity 40.3%; Pred. No. 1.2e-32;
 Matches 149; Conservative 52; Mismatches 130; Indels 39; Gaps 15;
 QY 695 VAAVDAAGNTSAPSTPYDCTPPGNQ-----GVTSDGGEYRQVTENMSSAQCCLTINT 749
 DB 16 1a1l1a1a11taapag-----anqgldrygtltlqg-ryvvgnmrwgtatqgc--lnv 65
 QY 750 AAGCAWVTSTANFSGGTGAPATYPSITKGCWGNCTTKNVGMPIQISQISAVTSMTTO 809
 DB 66 tng1gf1t1q1ad1q1d1s1v1plng1apks1ydv1qch1ync1ap1rtt-1pm1rl1s1s1g1s1a1p1s1v1ry 124
 QY 810 VSSGADVAYDIDWNTSTPTTGGOPNGTEIMTWLNSRGVOPFGSQRTGTVAGHTWNNW 869
 DB 125 tng1gvyuaeyd1wldptptng-vnrteim1wfnrvpvgp1gsyvgt-ahvgsrsewvrgs 182
 QY 870 QGQOTSKIISYVLTFGATISINLDKATFADAAAGSLNTSD-YLDDVEAGFEIWOGGO 928
 DB 183 tng1gndv1sf-1apsalswsvfvdv-fvdqavshglapdwyl1ta1sgfepwge9t 240
 QY 929 GLGSNSFSVSVTSQG-----VACRAITYVNSDMGSGFATYATVINTGSRAVNGTWV 979

```

Db      241 |glavnsfssavnagngggtpgtpacqvstsh-wpggfvtvdtlhtnigtbgdswel 299
        ||||| : : ||| || : : ||| ||||| : || :
Oy      980 AMSFGNQTVINRYMTALTSGASTATNLKSN-NVIOGOSTITGFNGSYSGTN-AAPT 1037
        :: |||| : || : : ||| : || : || : ||| : ||| ||
Db      300 dftlpaghtvesawmlispasgavlarslgsgilaangtfgsfqgtsqgfnapa 359
Oy      1038 -----LSTC 1041
        |||
Db      360 ggrlngtact 369

RESULT   8
AAB14880
AC AAB14880 standard; Protein: 371 AA.
XX
XX AAB14880;
XX
XX 21-NOV-2000 (first entry)
DE Streptomyces lividans CelB EGIII-like cellulase.
XX
XX Streptomyces lividans; Trichoderma reesei; endoglucanase III; EGIII;
KM cellulase; mutant; enzyme stability; textile treatment;
KW wood pulp treatment; feed additive; detergent.
XX
OS Streptomyces lividans.
PN MO200037614-A2.
PD 29-JUN-2000.
XX
XX 12-NOV-1999; 99WO-US26704.
XX PF
XX PR 18-DEC-1998; 98US-0216295.
PA (GENEV ) GENENCOR INT INC.
XX
XX Mitchinson C, Wendt DJ;
PI
DR WPI; 2000-482483/42.
XX
XX Novel endoglucanase III or endoglucanase III-like cellulase useful for
PT treating textiles and wood pulp comprises a substitution or deletion at
PT specified positions in the wild form of endoglucanase III -
PT
XX Example 1; Fig 3; 52pp; English.
XX
XX The present sequence is a cellulase related to endoglucanase III (EGIII)
CC from Trichoderma reesei. EGIII-like genes were isolated from genomic DNA
CC libraries constructed from various microorganisms by PCR. The isolated
CC genes showed significant homology to EGIII from T. reesei. Certain
CC substitution and deletion mutations have been incorporated into EGIII and
CC EGIII-like cellulases to produce variant enzymes with improved stability,
CC e.g. increased resistance to temperature stress. The mutants may be used
CC in textile and wood pulp treatment, as a feed additive, and for reducing
CC biomass to glucose. They are also useful for stonewashing or indigo dyed
CC denim and as an agent in laundry and dish detergents.
CC
CC Sequence 371 AA;
SQ

Query Match 11.5%; Score 635.5; DB 21; Length 371;
Best Local Similarity 40.3%; Pred. NO. 1.2e-32;
Matches 149; Conservative 52; Mismatches 130; Indels 39; Gaps 15

695 VAAVDAGNTSAPSTPVCTPGPNON-----GVTSVODEYRVQTNENWSSAQOCITINT 749
::: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 16 laslgaalltaaeapq-----angqidrygtlttldg-ryvvgnmrwgtstgc--lhv 65

750 ATGATVTSTANFSGGTGCAPATPYSIRYKGCHMGNCCTTKNVGMPIQISQIGSAVTSWSITQQ 809
::: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 66 tgnqfelcpadgsvglncgpkspkydvqgchycncaprll-lpmklsispspsvsyxy 124

```

QY	810	VSAAVYADVADIMWNSPPTTGGPNCGETIMWMLNSGSGVQPPGSGQATGACVTVAGHTMVMW	869
Db	125	tqngvnaayolwldprrtng-vntetlmhltfnrvpwpqipspgvt-ahvggtswevw	182
QY	870	QGQGTSMKRIISVYVTPCATSISMLDKAIFADAAANGSINTSD-YLLDVEAGFEIIMGQ	928
Db	183	tqngsnndvlsf-lapalsawefdkd-lvdqavshglatpdkwyltsdqgfepeggt	240
QY	929	GLGSNFSVSYSVTSGC-----VACRATYYVNSDMWSGGFATVTVTNTGSRATNGMTV	979
Db	241	glavnsfssaaavgngqfpgtgaacqysvsh-t-wpgsfvtdtltnsgtpeydwel	299
QY	980	AMWSGGNGOTVNTWNTMLTQSGASVYATNLKSN-NTIQGGQSTTRGFNSYSGTN-AAPT	1037
Db	300	dflpaghtrvssamwllspasgavatastcngriaangtqsgtfgtqssgtgfna	359
QY	1038	-----LSCRT 1041	
Db	360	ggrlntgstc 369	
RESULT	9		
AA	AY84345		
XX	AAAY84345	standard; Protein; 371 AA.	
XX	AAAY84345:		
XX	12-JUL-2000	(first entry)	
XX			
XX		Amino acid sequence of an endoglucanase III (EgIII)-like cellulase.	
DE			
XX			
XX		Endoglucanase III; EgIII; EgIII-like cellulase; surfactant stability;	
KW		cellulase; textile processing; textile cleaning; stonewashing;	
KW		indigo dyed denim; cellulose containing fabric; fabric smoothness;	
KW		pill removal; fibril removal; cotton; cellulosic fibre; dyeing; detergent;	
KW		animal feed; wood pulp; paper; grain; biomass reduction; glucose.	
OS			
XX		Actinomyces sp.	
XX			
XX		WO200014208-A1.	
PN			
PD		16-MAR-2000.	
XX			
PE		24-AUG-1999; 99WO-US19154.	
XX			
PR		03-SEP-1998; 98US-0146729.	
XX			
PA		(GEMV) GENENCOR INT INC.	
XX			
PI		Fowler T;	
DR			
XX		WPI; 2000-271052/23.	
PT			
PT		Novel variant endoglucanase III-like cellulases with improved	
PT		surfactant stability and resistance to temperature stress, useful for	
PT		textile processing or cleaning, treating wood pulp, food and grain, and	
PT		reducing biomass to glucose	
XX			
XX			
PS		Disclosure; Page 64-65; 73pp; English.	
XX			
CC		The present sequence represents an endoglucanase III (EgIII)-like	
CC		cellulase. The cellulase has homology to the Trichoderma reesei EgIII	
CC		protein. The variant cellulases have improved temperature stability,	
CC		and improved surfactant stability. The variant cellulases and	
CC		compositions containing them are used in textile processing or cleaning,	
CC		e.g. stonewashing of indigo dyed denim, and modifying the texture, feel	
CC		or appearance of cellulose containing fabrics (e.g. improving fabric	
CC		smoothness or removing pills and fibrils). The compositions may also be	
CC		used for the removal of immature or dead cotton from cellulosic fibres	
CC		or fabric, which can cause uneven dyeing. The cellulase may also be used	
CC		in a detergent composition for washing laundry and dishes and in the	
CC		treatment of animal feed, wood pulp, paper, non-animal foods and grains.	
CC			

CC The enzymes may also be used in the reduction of biomass to glucose.
 XX
 XX Sequence 371 AA:

Query Match 11.5%; Score 635.5; DB 21; Length 371;
 Best Local Similarity 40.3%; Pred. No. 1.2e-32;
 Matches 149; Conservative 52; Mismatches 130; Indels 39; Gaps 15;

QY 695 VAAVDAAGNTSAPSTPVDCTPFGNPNON-----GVTSVQDGEYRQNTNEMNSAQQCLTINT 749
 DB 16 laasgaallaaapq-----anqldcdygtltlqdt-ryvvnqmrvgtsatqc--lnv 65
 QY 750 ATGAMVTANFSGGTGAPATYPSIYKGCWNCCTTKNVGMFIQISOIGSAVTSMTQ 809
 DB 66 tngfcltqdgavspntgapskyspsvdygchyncaprt-llmrtsaigsapsvsyry 124
 QY 810 VSSGAIIVATYDWTNSTPTTTCGPNGTETIMILNSRCGVQPFSGOTATGVTYAGHTWTV 869
 DB 125 tngvynaaydlwldprtrng-vntelmlwfrvrgvpqlgspvgt-ahvggrawevw 182
 QY 870 OGQOTSMKIIISYVLTGATISNLDLKAIFADAAAGSLNTSD-YLLDVEAGEFIWOGQ 928
 DB 183 tsangendvstf-lapsaiswefdvkd-fvdqavshglapdvyltslqagfepwsgt 240
 QY 929 GLGSNSFSVTSVSG-----VACRATYVNSDMGSGFTATVYVNTGSRATNGTV 979
 DB 241 glavnsfssavnaagngngltpgtpaacqvsygtc-wpggtfvdtflntgspdvqwe 299
 QY 980 AMSFGNQTYNTYNTALTQSGASVATATNLYSN-NVIOGOSTTGFSGSYSGTN-AAPT 1037
 DB 300 dftlpgnhvtrtsaenallapsagavlarzsgngraingrlqsgtqsgtsgfna 359
 QY 1038 -----LSC 1041
 DB 360 ggrlmgfct 369

RESULT 10
 ID AAY84346 standard; Protein; 429 AA.
 XX
 AC AAY84346;
 XX
 DT 12-JUL-2000 (first entry)
 XX
 DE Amino acid sequence of an endoglucanase III (EGIII)-like cellulase.
 XX
 KW Endoglucanase III; EGIII; EGIII-like cellulase; surfactant stability;
 KW cellulase; textile processing; textile cleaning; stonewashing;
 KW indigo dyed denim; cellulose containing fabric; fabric smoothness;
 KW pill removal; fibril removal; cotton; cellulosic fibre; dyeing; detergent;
 KW animal feed; wood pulp; paper; grain; biomass reduction; glucose.
 XX
 OS Streptomyces lividans.
 XX
 PN W0200014208-A1.
 XX
 PD 16-MAR-2000.
 XX
 PF 24-AUG-1999; 99WO-0519154.
 XX
 PR 03-SEP-1998; 98US-0146729.
 XX
 PA (GEMV) GENENCOR INT INC.
 XX
 PI Fowler T;
 XX
 DR WPI; 2000-271052/23.
 XX
 PT Novel variant endoglucanase III-like cellulases with improved
 PT surfactant stability and resistance to temperature stress, useful for
 PT textile processing or cleaning, treating wood pulp, food and grain, and

PT reducing biomass to glucose
 XX
 PS Disclosure; page 65-66; 73pp; English.
 XX

CC The present sequence represents an endoglucanase III (EGIII)-like
 CC cellulase. The cellulase has homology to the Trichoderma reesei EGIII
 CC protein. The variant cellulases have improved temperature stability,
 CC and improved surfactant stability. The variant cellulases and
 CC compositions containing them are used in textile processing or cleaning,
 CC e.g. stonewashing of indigo dyed denim, and modifying the texture, feel
 CC or appearance of cellulose containing fabrics (e.g. improving fabric
 CC smoothness or removing pills and fibrils). The compositions may also be
 CC used for the removal of immature or dead cotton from cellulosic fibres
 CC or fabric, which can cause uneven dyeing. The cellulase may also be used
 CC in a detergent composition for washing laundry and dishes and in the
 CC treatment of animal feed, wood pulp, paper, non-animal foods and grains.
 CC The enzymes may also be used in the reduction of biomass to glucose.

XX Sequence 429 AA:

Query Match 11.5%; Score 635; DB 21; Length 429;
 Best Local Similarity 39.5%; Pred. No. 1.5e-32;
 Matches 152; Conservative 45; Mismatches 144; Indels 44; Gaps 12;

QY 675 PVTSTPTDGLAAGTAYTYTYAANDAGNTSAPSTPVDCTPFGNNGVTSVQDGEYRQVT 734
 DB 69 pqcvlatdgy-----frvtqdgasapungapksyp-----svfngchytq 109
 QY 735 NEMNSAQQCLTINTATGAMVTANFSGGTGAPATYPSIYKGCWNCCTTKNVGMFIQ 794
 DB 110 ntwgsatapcvta-ldtg-frvtqdgasapungapksypvfyngchync-spgtdlpr 166
 QY 795 ISOIGSAVTSWSTTOYSSAIVDAIDWTNSTPTTTCGPNGTETIMILNSRCGVQPFSG 854
 DB 167 ldtvsapsaisygfvdgaavnaaydlwldparcdg-vntelmlwfrvrgvpqlgsp 225
 QY 855 TATGVTYAGHTWVWMOGQOTSMKIIISYVLTGATISNLDLKAIFADAAAGSLNTSDYL 914
 DB 226 vgt-asvgrtewswsgngsndvlsfv-apsaiswefdvndfvratvarglaendyl 283
 QY 915 LDVEAGEFIWOGQGLGSNSFSVS-----TSGGV-----ACRATYVNSDMGSGFTA 962
 DB 284 tsvagfepwngngqlavnsfslvetgtpgtdpgdpgspaacavysltuv-wqdgfta 342
 QY 963 TTYVTNTGSRATNGTVAAWSFGNQTYNTYNTALTQSGASVATATNLYSN-NVIOGOSTT 1022
 DB 343 dvtvlnvgtlapvdwqlafclpssqgrltanwasltpssgsvlatgashariapgsals 402
 QY 1023 GFNGSYSGTNAAPT-----LSC 1041
 DB 403 fgfgtygafaeptgtrngfct 427

RESULT 11
 ID AAY08473 standard; Protein; 371 AA.
 XX
 AC AAY08473;
 XX
 DT 28-JUL-1999 (first entry)
 XX
 DE Actinomycete sp. 36KD cellulase protein.
 XX
 KW Cellulase; detergent; animal feed; nutritional value; textile;
 KW stone washing; texture modification; appearance; cellulosic fabric;
 KW pulp; draining; paper; baking additive; starch treatment; grain;
 KW high-fructose corn syrup production; ethanol production; fibre reduction;
 KW milling.
 XX
 OS Actinomycetes sp.
 XX
 PN W09925846-A2.


```

OY 1038 -----LSCT 1041
      |||
Db 360 garlngtact 369

RESULT 13
AA84797
ID AAY84797 standard; Protein; 395 AA.
XX
XX AAY84797;
XX
XX 08-AUG-2000 (first entry)
XX
XX Amino acid sequence of a C1 endoglucanase polypeptide.
XX
XX C1 endoglucanase; promoter; mutant; expression-regulating region;
XX secretion signal sequence; enzyme; cellulase; xylanase; pectinase;
XX lipase; protease.
XX
XX Chrysosporium lucknowense.
XX
XX Key Location/Qualifiers
XX Peptide 1..28
XX /note="signal peptide"
XX
XX WO200020555-A2.
XX
XX 13-APR-2000.
XX
XX 06-OCT-1999; 99WO-NI00618.
XX
XX 06-OCT-1998; 98WO-EP06496.
XX
XX (AARL-) AARL INC.
XX
XX Emalfar MA, Burlingame RP, Olson PT, Sinttyn AD, Parriche M,
XX Bousson JC, Pynnonen CM, Punt PJ, Van zelJl CMJ;
XX
XX WPI: 2000-317705/27.
XX
XX N-PSDB; AAA14818.
XX
XX Mutant Chrysosporium strain comprising nucleic acid encoding proteins,
XX particularly industrially important enzymes such as cellulases,
XX xylanases, pectinases, lipases and proteases -
XX
XX Disclosure: Page 43-46; 92pp; English.
XX
XX The present sequence represents a Chrysosporium C1 endoglucanase. The
XX polynucleotide sequence comprises a promoter sequence, and can be
XX linked to a nucleic acid sequence of the invention to control
XX expression. The specification describes a mutant Chrysosporium
XX strain which comprises a nucleic acid sequence encoding a polypeptide
XX of interest, where the nucleic acid sequence is linked to an
XX expression-regulating region and a secretion signal sequence.
XX The mutant strain is useful for producing heterologous proteins
XX and polypeptides of interest, e.g. industrially useful enzymes such
XX as cellulases, xylanases, pectinases, lipases and proteases.
XX
XX Sequence 395 AA:

Query Match 10.4%; Score 574.5; DB 21; Length 395;
Best Local Similarity 34.7%; Pred. No. 1e-28;
Matches 160; Conservative 58; Mismatches 156; Indels 87; Gaps 16;

OY 15 RGTTRMPAISKRLRAGYLAGVSIASIVPLAQHPAIAATHVDNPPYAGATFFVNPYMAQ 74
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 21 rttppkp-----rqsag-casavtldae-----tnvfqy---tlhpnntfyra 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 75 EYQSEANQNTATLAAKMRVSTYSTMMDRIAINGVNGRGOLTYTTLDAALSQOQGT 134
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 evaaseaaisdsaaieakrvadvgtflwidtleniy-----rllepal--edvpc 108
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

OY 135 PEVIEIYIDLPGRDCAALASNGELPATAQLQYETQYIDPLIASILSNPRYSLRIVTI 194
      :: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 109 enlyglvlydldprdcasakangel--kvgeIdtrYkIeyIdkIaeIlk--ahantafalv 164
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 195 IEPDSLNPNAVITNMSIQACATAVPYIEOGIEYALTFLKLAIPNVITYYMAASGMLGPNNA 254
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165 IepdsIphIvtnsdlqtcgsaasgyregvayalKqLn-IpvnvyIdagIhgyvIdanI 223
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 255 SGYQVEQVYLNASIGVNGIDGFEVTNTANTPTLKEPMTATQOQGGOPVESANFYQNNPD 314
      ||: | :: : : | | | : | | | : | | | : | | | : | | | : | | |
Db 224 kpgaqelaasyksaagspsqyrgIstnvaqw-----nawdqeprge-fdsadaqynkc 274
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 315 IDEADYAVDLYSRLVAAFPSSIGMLIDTJLNGMGPNPTGPSTAVDVNTFVNQSKIDL 374
      : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 275 qnekIyIntftgaelksaagmnh--aIdtgrngvtg-----1 309
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 375 ROHRGLMCNONGACLGOPPOASPTDFPNAHLDAVYWKIPQGESGTSNASDPTTGKSDP 434
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 310 rdewgdwcnvngagfyvprantgd---eladaftvwkpgsgsdgtads----aaryds 362
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 435 MCDPTTTSYGVLTNALPNSPIAGQWFPAPQDOLVANARPA 475
Db 363 fcgk-----pdafkpspeagwtmqayfemIlknanps 394
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
ID AAY01077 standard; Protein; 473 AA.
XX
XX AAY01077;
XX
XX 08-JUN-1999 (first entry)
XX
XX H. insolens Cel6A fungal cellulase protien sequence.
XX
XX Cel6B; Cel6A; fungal cellulase; cleaning composition; conditioner;
XX cellulolytically active protein; endo-1,4-beta-glucanase; enzyme;
XX fabric softener; Humicola-like cellulase; glycosyl hydrolase family 6;
XX detergent composition.
XX
XX Humicola insolens.
XX
XX WO9901544-A1.
XX
XX 14-JAN-1999.
XX
XX 02-JUL-1998; 98WO-DK00299.
XX
XX 04-JUL-1997; 97DK-0000813.
XX
XX (NOVO ) NOVO-NORDISK AS.
XX
XX Andersen KV, Damgaard B, Lund H, Nielsen JB, Schulein M;
XX
XX WPI: 1999-106046/09.
XX
XX N-PSDB; AAX27948, AAX27958.
XX
XX Cleaning composition containing Humicola endo-beta-1,4-glucanase
XX useful as detergent compositions or additives, or as fabric
XX conditioners
XX
XX Example 3; Page 243-245; 271pp; English.
XX
XX This sequence is the Humicola insolens Cel6A fungal cellulase.
XX The invention relates to a cleaning composition (A) that contains at
XX least one enzyme (I) with cellulolytic activity and at least 25 wt.% of
XX total cellulolytically active protein present is Humicola
XX endo-1,4-beta-glucanase or a Humicola-like cellulase of the glycosyl
XX hydrolase family 6. (A) are used as detergent compositions or additives,
XX or as fabric softener or conditioner. (I) provides colour clarification,
XX and possibly soil removal, without significant weakening of cellulostic
XX materials during pre-soaking or wet storage (contrast use of enzymes from

```


Db 407 mriptant---ghqyvdafrwvkkpggcdgts----dttaarydyhcg-----leda 450
Oy 451 LPNSPIAGOWFPQAQFDQLVANARP 474
| : | | | | | : | | |
Db 451 lkpaapeagqwfqayfegqlrnanp 474

Search completed: August 29, 2002, 16:21:46
Job time: 321 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:24:05 ; Search time 72.78 Seconds
(without alignments)
1377.043 Million cell updates/sec

Title: US-09-917-384-6
Perfect score: 5536
Sequence: 1 MERTQSGRCRCRQGRITRM.....GFGSYSGTMAATLSTAS 1043

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1707	30.8	872	2 S49541	cellulase - Cellul
2	1239.5	22.4	683	2 A82704	1,4-beta-cellobios
3	1051	19.0	579	2 T35240	probable secreted
4	1000	18.1	596	2 A55976	cellulose 1,4-beta
5	641	11.6	438	1 S70602	cellulose 1,4-beta
6	629.5	11.4	382	2 JC2571	cellulase (EC 3.2.
7	586	10.6	1090	2 S59077	cellulose 1,4-beta
8	561.5	10.1	1045	2 A39199	endoglucanase B (E
9	548	9.9	471	1 A26160	cellulose 1,4-beta
10	548	9.9	471	1 A38979	cellulase (EC 3.2.
11	487	8.8	747	2 B47053	cellulose 1,4-beta
12	450.5	8.1	1711	2 T31337	cellulase (EC 3.2.
13	449.5	7.8	1742	2 T17120	cellulase (EC 3.2.
14	434	7.8	1331	2 A48954	cellulase (EC 3.2.
15	431.5	7.5	1779	2 T31085	cellulase (EC 3.2.
16	413	7.0	1854	2 S36859	cellulase (EC 3.2.
17	388.5	6.9	700	2 B41897	cellulase (EC 3.2.
18	384.5	6.9	890	2 T35237	cellulase (EC 3.2.
19	383.5	6.9	1039	2 S02711	cellulase (EC 3.2.
20	383	6.9	498	2 A27196	cellulase (EC 3.2.
21	382	6.9	508	2 G65993	cellulase (EC 3.2.
22	381.5	6.9	499	2 JN0111	cellulase (EC 3.2.
23	380	6.9	145	2 A41897	cellulase (EC 3.2.
24	379	6.8	504	2 A26874	cellulase (EC 3.2.
25	363.5	6.6	504	2 S54744	cellulase (EC 3.2.
26	363.5	6.4	380	2 G70847	probable cellulase
27	355.5	6.3	505	2 S39962	endoglucanase - E
28	351.5	6.3	1162	2 T30433	scaffolding protei
29	351	6.3	915	2 A43802	cellulase (EC 3.2.

30	349	6.3	586	2 PC6006	scaffolding protei
31	339	6.1	1483	2 C97012	probably celluloso
32	338.5	6.0	441	2 T12011	cellulase (EC 3.2.
33	333	6.0	486	2 I40548	bifunctional cellul
34	325.5	5.9	986	2 S12021	thermoactive cellul
35	316	5.7	756	2 AB1452	chitinase B homolo
36	315	5.7	699	2 A38368	chitinase (EC 3.2.
37	310	5.6	756	2 AB1088	chitinase B homolo
38	309	5.6	1848	2 A41410	cellulase-binding
39	299	5.4	426	2 A42360	cellulase (EC 3.2.
40	289	5.2	879	2 A47704	cellulase (EC 3.2.
41	285.5	5.2	570	2 S56132	endoglucanase I (E
42	277.5	5.0	5188	2 B85547	probable RTX famil
43	277.5	5.0	5291	2 F90696	hypothetical prote
44	276	5.0	1230	2 S47466	cellulose 1,4-beta
45	275	5.0	488	2 A41961	chitinase (EC 3.2.

ALIGNMENTS

RESULT 1
S49541
Cellulase - Cellulomonas flm1
C:Species: Cellulomonas flm1
C:Date: 29-Nov-1995 #sequence_revision 01-Mar-1996 #text_change 22-Oct-1999
C:Accession: S49541; A47093
R:Meinke, A.; Gilkes, N.R.; Kwan, E.; Kilburn, D.G.; Warren, R.A.J.; Miller Jr., R.C.
Mol. Microbiol. 12, 413-422, 1994
A:Title: Cellulose-binding polypeptides from Cellulomonas flm1: endoglucanase D (Cend
A:Reference number: S49541; MUID:94344030
A:Accession: S49541
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-872 <ME2>
A:Cross-references: EMBL:J25809; NID:9456028; PIDN:AC36898.1; PID:9456029
J:Meinke, A.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.
B: Bacteriol. 175, 1910-1918, 1993
A:Title: Cellulose-binding polypeptides from Cellulomonas flm1: endoglucanase D (Cend
A:Reference number: A47093; MUID:93209933
A:Accession: A47093
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 812-872 <ME2>
A:Experimental source: ATCC 484
A:Note: sequence extracted from NCBI backbone (NCBIN:128120, NCBIPI:128121)
F:769-870/Domain: bacterial cellulose-binding domain homology <BCB>
F:770-869/Disulfide bonds: #status Predicted

Query Match	30.8%	Score 1707;	DB 2;	Length 872;
Best Local Similarity	38.6%	Pred. No. 1.4e-84;		
Matches 403;	Conservative 123;	Mismatches 311;	Indels 206;	Gaps 27;
OY	23	ISKRRACVLAGAVSIASTVPLA--MOHPAIAATHYDNYACATFVNPVNAOVSER 80		
DB	10	VARRRRA-VATRAATFALVAAPLTATASAAVHVDNPPAGVQVYVNTMAASVMAA 68		
OY	81	ANCT-NATLAAMRVSTYTAVMMDRIAINGVNGSGGLTYDAALSOQOCT-TPEVI 138		
DB	69	GROSDAPALAAAKRTVAGOPTVAMMDRISALTGNADSGLEFHDNNAVADOKAGVPLV 128		
OY	139	EIVYDLPGRDCAALASNGELPATAGCLOTETOTYIDPAISLSPNYSRIYTLIEPD 198		
DB	129	NATVYDLPGRDCAALASNGELPATAGCLOTETOTYIDPAISLSPNYSRIYTLIEPD 188		
OY	199	SLPNATVNSIOMCAFAVYEEGIEYATLTKLAIPNYIYMDAAHSGWLGPNNK---S 255		
DB	189	SLPLTLTINISEPCQQAAPYRKGAVYALDKHAIPVNYNYIDIGHSQWLMDSNACPSA 248		
OY	256	GYVOEVQKVLNASIGVNGIDFVNTANTYPLKEPMT-ATQVVGQVESANFYQWNP 314		

Db 249 TLEAEVAK--STTAGFASIDGFVSDVANTTPLEEPPLSSSLTINTPIRSSKFEYEMND 306
 Qy 315 IDEADYAVDLVSRVLAAGPSSIGMLIDTLRNGWCGNPEPTGPSTATDVNTEFVNSKIDL 374
 Db 307 FDEIDYTHMHLRLVLAAGPSSIGMLVDTSRNGWCGNPNPSTITASTDVNAVYADNRRVD 366
 Qy 375 RQHRGLMGNONAGGQPPQASPTDFPNHLDAYWIKRPGSGSDGTSASDPTTKKSP 434
 Db 367 RYHRKAMCPLGAGIGRPEAPPSYASHDLAEFWIKRPGSGSDASDIPNDQKRFDR 426
 Qy 435 MCDPFTYTS--YGVLTNMLPNSPIGOWEPAOFDOLVANARAVVSGGLKYKKNDSAP 492
 Db 427 MCDPFTYTSKLNQTLGATPNAPLAGOWEPEGFVLTKNAYVI--GG-----TTP 475
 Qy 493 GDNQKRP-----GLQVNTGSSVDLSVTVRYWFTRDGSSSTLVYNCDMAMCGNIRAS 548
 Db 476 VEDLAPVPVPGLTAGTTATATSVPLS-----WTASTDVAVATGYGVYRGTLVGTAAAT 529
 Qy 549 FGSVNPAPPTADTYQLSFTSGTTLAAGSGTGEIQNRVKNKSDMSNTEFTDYSYGTNTTQ 608
 Db 530 SYTVTGLTP-ATAY--SFTYRAKADAGNVSA-----SAAAAATTQ 567
 Qy 609 DWTKYTVYVNGVLVWGTEDVPPSVPTGLVTVGSSSVSLAMNASTDNV---GVAHYNV 665
 Db 568 SGT-----YDTTAPSVPAGLAGTTTTPPLSMWTASTDAGSGSVAGYEV 614
 Qy 666 YRNGVLVGPVYTSFTDGLAGTAYTYVAADVAGNTSAPSTPYDCTPGPNONGVTSV 725
 Db 615 LMGTVVGTATSYTVGLTGTGYTSFSVRAKDAVAGNTSASAASVAT----- 663
 Qy 726 QDGEYRVQCTENENSSAQOCLTINTAGMTVSTANPSGSGTGAPATYPSIKYCHMGNT 785
 Db 664 -----TOTGYVDTTA-----PSVPGTLAGTTT 687
 Qy 786 TKNVGMPQISQISGAVTSMSTTOVSSGAYDVA-YDIWNTSPPTTGOPNGTEIMIMLNS 844
 Db 688 TSSVPL-----TWASTDNAGSGVAGYEVNGTTRVAT----- 721
 Qy 845 RCGVOPFGSQATGYTVAGHTNNWQGOOTSMKIIISYVLTLPATISLNDLKAIFADAAA 904
 Db 722 -----VTSSTYTVG-----LAADTAY 738
 Qy 905 RGSLSMTDYLLEVEAGFEIMOGGGLGNSFSVSYTSGGVACRATYVYVNDMSGSTAY 964
 Db 739 SFTYVAKDVAGNVSA-----SAANSARTQAATSGS--CTVYKISAS--NNTGTTGTV 787
 Qy 965 TWTNTGSRATNGTYVAAWSPGNGQTYTNWNTALTQSGASVTATNLSNNVIOPGOSTTTP 1024
 Db 788 EYKNNGTALNGLTIGFSPADGQKVSQGSAMESQSGTAVTAKNAPWNGTLAGSSVSIG 847
 Qy 1025 FNGSYSGTNAATP-----LSCT 1041
 Db 848 FNGTHNGTNTAPTATLNGVACT 870
 RESULT 2
 A82704
 1.4-beta-cellobiosidase XFI267 [Imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: A82704
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717
 A>Note: for a complete list of authors see reference number A59328 below
 A:Accession: A82704
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-683 <SIN>
 A:Cross-references: GB:AE003960; GB:AE003849; NID:g9106242; PIDN:AA84076.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carier
 as-Neio, E.; Docena, C.; El-Dorri, H.; Facinani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 J:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
 J.D.; Junqueira, M.L.; Kemper, E.L.; Klajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
 A:Authors: Martins, E.M.F.; Matsukuma, A.T.; Menck, C.R.M.; Miracca, E.C.; Miyaki, C.
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
 M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XFI267

Query Match 22.4%; Score 1239.5; DB 2; Length 683;

Best Local Similarity 31.9%; Pred. No. 1.8e-59; Mismatches 314; Conservative 124; Mismatches 220; Indels 327; Gaps 27;

Qy 53 AATYDNPYAGATPFVNPVMAOEVOSEANQTNATLAAMRVSYSTAYVMDRIATANG 112
 Db 18 AEAHNDNPFVADATSYLNDPYSKIDISTIAKNDVSLKEMQITKPYGVWIDITDAKVG 77
 Qy 113 VNGCPG---LTTYDLAALSQOQGTPEVEIYIYDLPGRCAALASNGELPATAAGLOTY 169
 Db 78 GPKNGRLMLTGLDAAALAKGKNTPITATFVIYDIPGDCHALASNGELPLTPGLORY 137
 Qy 170 ETOYIDPIASLNSPKYSRLVITIEPDSLPNATYNNNSIOCATN--VPYEGDIEVAL 227
 Db 138 KREYIDTASISNPKYKIDIRIVNIEPDSLPIITNNSTPSCAOKAKNTGYEBSIKYAL 197
 Qy 228 TKLHAIPNVYIYMDAHSGLMGPNNASGYVOEVKVL-NASIGYNGIDGFEYNTANTYTP 286
 Db 198 NKLSEIPNYYNMDIGHSGMLGMDNRIIPAVSLYKTRIGSTAGASVNGFINTANTTP 257
 Qy 287 LKEPRN-TATQVGGQPPESANFYQWNPDIADYADVLSYSLVAAGPSSIGMLIDTLR 345
 Db 258 LIEPNLPNDLILIGQPIRSSKFEYENRYFEDMDSELYXNDFVAAWPSISGFIIDYGR 317
 Qy 346 NMGGPNPEPTGSTATDVTFVYNSKIDLRHGRGLMONGNGGLGLOPPASPTDPNNML 405
 Db 318 NMGGPERFTS-AFGNDVNSIYNSGRIDRRHGMCKDKDGIGIPAT---PGGHV 372
 Qy 406 DAYVWIKPPESDGTSAASDPYTGKSKSPDCPTTYTSGVLTNMLPNSPIAGOWEPAOF 465
 Db 373 DAFQWIKPPEYDSSSLIPNDQKGFEDYCDPTFTTPGVLTGALPAPLSGDFHNOF 432
 Qy 466 DOLVANARPAVYSGGLKVOYKKNDSAPGDNQITKPLQLVNTGSSSVDSVTVRYWFTRD 525
 Db 433 VKLINNAYPDI-----SRTPLPG-----SGSSS-----S 456
 Qy 526 GGSSTLVYKNCMAAMAGCNIIRASFGSVNPATPTADTYQLSFTSGTTLAAGSGTGEIQNRV 585
 Db 457 SGSS-----SSSGS-----SSSGSSSSSSSS----- 479
 Qy 586 NKSMSNFTETNDYSYGTWTFQDMTKYTVYVNGVLVWGTEDVTPPSYTGVLVTVGSGS 645
 Db 480 -----SSGSSS-----SSSGSS-----SSSSSS 492
 Qy 646 SYSLAMNASTDNVGAHYVYVYVNGVLVGPVYTSFTDGLAAGTAYTYVAADVAGNTS 705
 Db 493 SGSSSSSSSSSSSG-----SSSGSSSSSSSS-----SSSSSSSSSS 515
 Qy 706 APSTPDCIPRGNONGVTSVQDGEYRVQCTENENSSAQOCLTINTAGMTVSTANPSG 765
 Db 516 SSGS-----SSSGSSSSSSSSSS-----SSSGSSSSSSSSSS 537
 Qy 766 GGAPATYPSIKYCHMGNTKRVGMPQISQISGAVTSMSTTOVSSGAYDVAVDIWTNS 825
 Db 538 S-----SSSGSSSSSSSSSS-----SSS 558

RESULT 3
 T35240
 probable secreted cellulase - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C:Accession: T35240
 R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, September 1998
 A:Reference number: Z15172
 A:Accession: T35240
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-579 <SEP>
 A:Cross-references: EMBL:AL011515; PIDN:CAA20645.1; GSPDB:GN00070; SCOEDB:SC5C7.33
 A:Experimental source: strain AS(2)
 A:Genetics:
 A:Gene: SCOEDB:SC5C7.33

[illegible][illegible]

```

RESULT      5
S70602
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) II precursor - cultivated mushroom
N:Alternate names: cellulase
C:Species: Agaricus bisporus (cultivated mushroom)
C:Date: 10-Sep-1999 #sequence_rev1sun 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: S70602
R:Yaguee, E.; Chow, C.M.; Challen, M.P.; Thurston, C.F.
Curr. Genet. 30, 56-61, 1996
A:Title: Correlation of exons with functional domains and folding regions in a cellul
A:Reference number: S70602; MUID:96269330
A:Accession: S70602
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-438 <YAG>

```

A:Cross-references: EMBL:Z34007; NID:g1494969; PIDN:CAA83971.1; PID:g1494970
 C:Genetics:
 A:introns: 34/1: 56/2: 96/3: 202/1: 325/3: 410/2
 C:Superfamily: cellulose 1,4-beta-cellulobiosidase II: fungal cellulose-binding domain hom
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:25-56/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 11.6%; Score 641; DB 1; Length 438;
 Best Local Similarity 33.0%; Pred. No. 2e-27;
 Matches 162; Conservative 64; Mismatches 145; Indels 92; Gaps 14;

14 QRGTRMPAISRRLRAGVAGAVSIASIVPLAMQHPAIAHVDNPVAGATFENPYWA 73
 DB 68 QPGTR-PRATTTSGGTGPTSGA-----GNPTGKTVMTSPFA 104
 QY 74 QEVQSAANQINATLAARKRVSTYAVWMDRIAINGVNGPGLTYYLDAALSOQOQT 133
 DB 105 DEVAQAAADISNPISLATKASVAKIPTFVWFDTAKV-----PDLGGYLAARSKNO-- 156
 QY 134 TPEVEIYIYDLPGRDCAALANGELPATAGLQTEYQYIDPIASILSNPKYSSLRITYT 193
 DB 157 --LVOIIVYDLPDRDCAALANGESLANDGLNRYK-NYDQIAAQIK--OEPDVSVA 210
 QY 194 IIEPDSLPAVNTNMSIQACATAPVYEOGIEYALTKLHAIPVYIYMDAAHSGWLGMPN 253
 DB 211 VIEPISLANLTNTLVQCKANQASAKREVITAYOKLNAV-GVTMYIDGHCWGLGMPN 269
 QY 254 ASGVQEVQKVLNASIGVNGIDGVTNTANTYPLKEPEMTATQOYGQOPVESANFYQNP 313
 DB 270 LSPAQLQFOIYRDAGSPNRLNATVNFNALRAS-----SPDIT-----QGSN 316
 QY 314 DIEDADYANDLSRLVAGFSPSIGMLIDTLNGMGCPNEPTGPRATVNTFVNSKID 373
 DB 317 NDETHYITALPMLSNAGEPAH--FTVDGRSG-----VGN 351
 QY 374 LRQHRGLMCNONGAGIGOPPOASPTDFPNAHDAVYIWKPESESQTSASDPPTGKKSD 433
 DB 352 INDQMDMCNVKAGAFGQRP---TNTGSSLDAIYMWKPGEGCDTSSNSSP-----RFD 404
 QY 434 PWCDFYTTTSTYVLTNALPNSPIQWMPAPQDQLVANRPV 476
 DB 405 SHCS-----LSDAHQPAPEAGTWQAFETLVANANPAL 438

RESULT 6

UC2571
 cellulase (EC 3.2.1.4) precursor - Streptomyces rochei (strain A2)
 N:Alternate names: endo-1,4-beta-glucanase; endoglucanase
 C:Species: Streptomyces rochei
 C:Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 22-Oct-1999
 C:Accession: J02571; S34392
 R:Perito, B.; Hanhart, E.; Irdani, T.; Iqbal, M.; McCarthy, A.J.; Mastromel, G.
 Gene 148, 119-124, 1994
 A:Title: Characterization and sequence analysis of a Streptomyces rochei A2 endoglucanase
 A:Reference number: J02571; MUID:95011642
 A:Accession: J02571
 A:Molecule type: DNA
 A:Residues: 1-382 <PER>
 A:Cross-references: EMBL:X73953; NID:g393391; PIDN:CAA52139.1; PID:g393392
 A>Note: this cellulolytic strain was isolated from the gut of termites
 C:Genetics:
 A:Gene: egls
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
 A:Pathway: cellulose degradation
 C:Superfamily: bacterial cellulose-binding domain homology
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:1-37/Domain: signal sequence #status predicted <SIG>
 F:38-382/Product: endoglucanase #status predicted <MAT>
 F:279-380/Domain: bacterial cellulose-binding domain homology <BCB>
 F:280-379/Disulfide bonds: #status predicted

Query Match 11.4%; Score 629.5; DB 2; Length 382;
 Best Local Similarity 39.9%; Pred. No. 7.1e-27;
 Matches 147; Conservative 48; Mismatches 148; Indels 25; Gaps 9;

685 LAAGTAVTYTVAADAGNTSAPSTPVDOCTPGPNONGVTSVODGEYRVQTNEMNSAQOC 744
 DB 11 LRAVSNALLITALLAALITATAPPAADTTI-CEEFSGTVID-GRYVQNNRWGTSATQC 68
 QY 745 LTINATGAMVSTANFSGGTGADATYPSIYKGCWNGCTTKYNGMPDIQISQASVTS 804
 DB 69 VT--ATDSGFRTQADGSPVTPNGAPKSPSVENCGHYTNC-SPGALPARISGISASRSS 125
 QY 805 WSTQYSSAIVAAIDITNSTPTTGGPNGCEIMIMLSKCGVOPFSGQATGTYVAGH 864
 DB 126 ISYGFVDNAVYASVDIADPPRTDQ-VNRTEIMWERNQOIQIPISQVGT-ASVAGR 183
 QY 865 TNNWQOQOQTSWKIISYVLTGATSIISNLDKATFADAAARSLNMTSDYLDVEGFETW 924
 DB 184 TWEVMSGNGTNDVLSFV-APSAMSSWSEFVMDYFVATVATGACNDWTLISIQGFEPW 242
 QY 925 QGGGIGSNSEFSVSTSG-----VACRAIYVNSDMGSGFTATVVTN 968
 DB 243 QNGAGIAYVFSFSTVNTGSSQNDPBNPBGDPFPAACVSVATNV-MPGGFTANVTN 301
 QY 969 TGSRAITNGTVAWSFGNQTNTNNTALTQSGAVTATNLXSNVIOPGOSTTGFGNS 1028
 DB 302 NSAPVDGWRILFTLPSSQSVYHANNASVSPSGAVTATGPAESARIAAGSGSQSFGQA 361
 QY 1029 YSGTNAAP 1036
 DB 362 YGSFRAQP 369

RESULT 7

SS9077
 cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) precursor - Cellulomonas fimi
 N:Alternate names: exo-beta-1,4-cellulohydrolase
 C:Species: Cellulomonas fimi
 C:Date: 15-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 22-Oct-1999
 C:Accession: S59077; S59044; I40694
 R:Shen, H.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.J.
 Biochem. J. 311, 67-74, 1995
 A:Title: Cellulohydrolase B, a second exo-cellulohydrolase from the cellulolytic ba
 A:Reference number: S59044; MUID:96003698
 A:Accession: S59077
 A:Molecule type: DNA
 A:Residues: 1-1090 <SHB>
 A:Cross-references: EMBL:L38827; NID:g790696; PIDN:AAB00822.1; PID:g790698
 A:Accession: S59044
 A:Molecule type: protein
 A:Residues: 54-78;456-461 <SHW>
 R:Shen, H.; Tomme, P.; Melnik, A.; Gilkes, N.R.; Kilburn, D.G.; Warren, R.A.; Miller,
 Biochem. Biophys. Res. Commun. 199, 1223-1228, 1994
 A:Title: Stereochemical course of hydrolysis catalysed by Cellulomonas fimi Cene, a m
 A:Reference number: I40694; MUID:94197708
 A:Accession: I40694
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 54-106 <RES>
 A:Cross-references: GB:L29042; NID:g556035; PIDN:AAA50257.1; PID:g556036
 C:Genetics:
 A:Gene: cbnB; cene
 C:Superfamily: bacterial cellulose-binding domain homology
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:1-53/Domain: signal sequence #status predicted <SIG>
 F:54-1090/Product: cellulose 1,4-beta-cellulobiosidase #status experimental <MAT>
 F:989-1090/Domain: bacterial cellulose-binding domain homology <BCB>
 F:990-1089/Disulfide bonds: #status predicted

Query Match 10.6%; Score 586; DB 2; Length 1090;
 Best Local Similarity 24.0%; Pred. No. 6.7e-24;

Matches 299; Conservative 147; Mismatches 419; Indels 382; Gaps 62;

20 MPALSKRLRAGLAVAGVSIASIVLMAQHPALATHVDPYAGATFFVVPYMAQVQSE 79
 1 MSSTRRRSAMAAATVGV-SSEFLAVAGITPAIAAGAGOP--AT-----VYTP 46
 80 AANQTNATLAAR--MRVSTSTAVMORIAINGVNGPGLTYLDALSOOGTTPREV 137
 47 AASPVAADVGEYAOEFLAOY-----DKIK--DPANG-----YFSA-----OGIPYHA 87
 138 IEIYIDLP--GRDCAALASN-----GELPTAAGL-QITEY-----OYDPIA 178
 88 VETLMEAPYGHETTSSEAYSWLMLEALYGVQVODMAPLNMHMTMKYIPQSVQDP 147
 179 SILSNPKYSSLRITVITIEPSPLEPNATN-----MSIOCATVAPYEGIEYALTKL 230
 148 NSFYNPSPATYVAFEFNHPSSYPSOLNSGISTGTDIGELKAT-----YGNADVOMHL 203
 231 HAIPRVYIYMDAHSQ-WLGMPPNNAAGYVOEOKVLNLSI-----GVNGI 274
 204 ADVDNITFGFATPGAGCTIGTATGTFINTFORQDESWETVPOPSCEEFKYGKNGY 263
 275 DEFVTNATYPLKEPEFATQOYGGGPVSS--ANFYQNPDPIDEA-----318
 264 LDFETKDAISA--KQKVTYSASDADARAEVAYWAN--QWATEQKADVAATVAKAAM 319
 319 -DY-----AVDLSRLV-----NAGFPSSIGMLIDILRNMGVPPNE-----353
 320 GDLRLTLEDKFKFKIGTSPCAAGOREAHYLLSYTMAMGATDTSSGMARIGSSH 379
 354 -----PTGPTATDVNTEFNQSKIDLROHRLKMCNONGI---389
 380 AHFGYONPLAALSTDPKLTSPKAKADMAAM--OROLEFY-----TWLOASMGILAG 434
 390 -----GQPPQASPTDFPNALHDAVYIKPPGES--DGTSA-----ASDPTT 428
 435 ATNSMCAVYQOPAGCTPTEFGMGYTAPEYVDPSPNRHMGAMQVORAEIYVSGNAQ 494
 429 GKSDPKDPTTYSYGVLTNALPNSFIAGQWPPAOFDOLVANARAVYSGGLKYOYKN 487
 495 AKK-----ILDKWVPVAVANLSTDGASWKPKPSELKTKGPD 530
 488 --NDSAPGDNDQIKPGLQI-VWFGSSVYDLSGVYVR--YWFTRDGGST-----LVYKCD 536
 531 TNNAAAPGCGN--PGLTYEYTSYGODVGAADTARALLFFAAKSGDPAKSKAKALDAI 587
 537 WAAAMCGNTRBSFGSVNPATPTADYIQLSFT-----GGLAAGSGTGEION-RVKKSPW 590
 588 WA-----NNODPLGV--SAVETRGDYKRFDDTYVANGDGIYIPSGWGTGTPNGDVIKPGV 640
 591 SNFTEENDSYGTNTTFODWTKYVYVNG-----VLWVG-----624
 641 SFIDISFYKKDPN-----MSKVOTFLDGGALPEQFRHRENAQTVAGALADVARLFDG 695
 625 --REDVTPSPVPTGLVVTGVGSSSVLANNASTDNVGAHVAVVYVNGVLVGOPTVSTFD 682
 696 TTTDTTAPVPTGIGAGVYTSFATISWTASTDTRVYGIVYGAIRKVGATATTSSTD 755
 683 TGLAAGTAYTYTVAADAAGTSAAPS-----TPVDTCPGPNONGYTSVQDEHYVQ 733
 756 TGLTFSTAYATVAFADAGVNSAPSAALVTTKATPSPD-TTAPSPATIS-----805
 734 TNNMSSAOCILITNTATGANTVSTANFSGGTGGAATPYSIKKGGHMCNCTTKNVGMI 793
 806 -----SSS-----TANSVYIGMSASTDN-AGGSG--LAGY-DVYRGA-----T 839
 794 QISQGSANTSWSTQVSSGAY-----DVAYDI-----WTNSTPTTGGQ--P 833
 840 RVAQT-TALTFTDGLITASTAYEYVRADVAGVNSAPSTAVSYTKSDTPTTAPSP 898
 834 NGTEIMWLNRSRGVQFSGQATGVTYAGHTWVWQOQTSWKLISVLTLPATISLNL 893
 899 AGLAAM-----TVEETSAV-LTWN-----ASTDYGGS 924

894 DLKA--IFADPAARGLNLTSDYL--LDVEAGEFE-----IMOGGCGIGSNSFSVSIS- 941
 925 GLKGYDYIRGATRVGSGTTSATSYDTGLTAATVAYOTVATRADAGVNSAASALSVTTKTP 984
 942 -GGYACATVYVNSDMGSGFTATVATVNTGSRATNGVYVNSFGGNGVYVNTVMTALQOS 1000
 985 QYSGSCSVAINASS--WMSGFTASVRIINCTTINMGSLGFDLTAGQKQVQCMKMTWQS 1043
 1001 GASVATNLVSNVYIOPGOSTTGFNGSYSGTNAAPT-----LSCT 1041
 1044 GSTVATNAPMNGTLPADQVTVDFGNSHTGONPNASFTLNASCT 1090

RESULT 8
 A39199
 endoglucanase B (EC 3.2.1.-) - Cellulomonas fimi
 C:Species: Cellulomonas fimi
 C:Date: 08-Nov-1991 #sequence_revision 08-Nov-1991 #text_change 22-Oct-1999
 C:Accession: A39199
 R:Meinke, A.; Braun, C.; Gilles, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.J.
 J. Bacteriol. 173, 308-314, 1991
 A:Title: Unusual sequence organization in CenB, an inverting endoglucanase from Cellu
 A:Reference number: A39199; MUID:91100298
 A:Accession: A39199
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1045 <MEI>
 A:Cross-references: GB:64644; NID:q144415; PID:AAA23086.1; PID:q144416
 C:Superfamily: fibronectin type III repeat homology; bacterial cellulose-binding doma
 C:Keywords: glycosidase; hydrolase
 E:945-1045/domain: bacterial cellulose-binding domain homology <RCB>
 E:946-1044/disulfide bonds: #status predicted

Query Match 10.1%; Score 561.5; DB 2; Length 1045;
 Best local similarity 26.7%; Pred. No. 1.3e-22;
 Matches 265; Conservative 105; Mismatches 382; Indels 241; Gaps 47;

146 PGRD-----CALASN-----GELPTAAGL-----QTEYQYIDPIASILSNP 184
 181 PGSDVAETRAALASAIYVILKDDPAVYATVSHAKOLYTRADYTRGASDCTAASAY 240
 185 K-YSSLRIYTIIPBDSLPMNAVNTMSIOCATVAPYEGIEYALTKLHAIPNYIYMDA 243
 241 KSMGQYDELWCAVYLYKATGATYLAKEA-EYDKLGTENQST-----TRS 287
 244 HSGWLGMPNNAAG-----YQDEVOKVINA-SIGVNGIDGFTVNTANYPLKE 289
 288 YKWTIANDNKOFGYVYALIAMETGOKRYVDANRWLDYTVGVNG-----QKVPYSP---338
 290 PPMATQOYGGQPY-----ESAN-----FYQWMPD-----IDEDYAV- 322
 339 -----GGQAVLDSMGLARVYANSTFVALVYSDMNTDAFRKARYHDFGROINYLALG 389
 323 -DLYSRLVYAGFPSSIGMLIDTLANGCGGPNPEPSTATADVNFVNOCKIDLROHRLGM 381
 390 DNPRESSYVVG-----GANPPLAHRHRAHSMWLSITTPAQSRLHYL 433
 382 CMONGAGIGQPPQASPTFPNHLDA--YVWIKPGEBSDCTSAASPTTGKXSDMCPD 438
 434 ---GALVGP--GSPND--AYTDSRDYV-----ANEVATDYNAG-----466
 439 TYTTSYGVLTNALPNSFIAGQWPPAOP-----FDOLVANARAVYSGGLKYOYKNDSAP 492
 467 -FTSALARLVEEXGGPPLASFPPEQPDGDLFVBAAMLNPPSGTFFEVKMINOGAPF 525
 493 GDNOIKPGLQLVWNGSSVDLSVTYVYVPTFDG--GSSTLYVNCWAMAGCGNTRBSFGS 551
 526 ARS-----LKNNAKRYWFTTIDGFAASDVTLSANYSE--CG--AAGSK 563
 552 VNPATPTADYIQLSFTGGLAAGST--GEIONRV-NKSDMSNFTENDYISG--TNT 605

RESULT 9
 A26160
 cellulose 1,4-beta-cellubiosidase (EC 3.2.1.91) II precursor - fungus (*Trichoderma reesei*)
 N:Alternate names: exo-cellulobiohydrolase II
 C:Species: *Trichoderma reesei*
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A26160; A26472
 R:Chen, C.M.; Grifzall, M.; Stafford, D.W.
 B:Bio/Technology 5, 274-278, 1987
 A>Title: Nucleotide sequence and deduced primary structure of cellulobiohydrolase II from
 A:Reference number: A26160
 A:Accession: A26160
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-471 <CHE>
 R:Theerl, T.T.; Lehtovirta, P.; Kauppinen, S.; Salovuori, I.; Knowles, J.
 Gene 51, 43-52, 1987
 A>Title: Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and
 A:Reference number: A26472; MUID:87248061
 A:Accession: A26472
 A:Molecule type: DNA
 A:Residues: 1-471 <TE>
 A:Cross-references: GB:M16190; NID:g170540; PIDN:AAA34210.1; PID:g170541
 C:Genetics:
 A:Introns: 31/2; 160/2; 243/1
 C:Superfamily: cellulose 1,4-beta-cellubiosidase II; fungal cellulose-binding domain hom
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:31-62/Domain: fungal cellulose-binding domain homology <FCB>

[illegible]

```

RESULT 10
A38979
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) II - fungus (Trichoderma viride)
C:Species: Trichoderma viride
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A38979
R:Wang, J.R.; Zhang, M.F.; Huang, T.
Acta Genet. Sin. 22, 74-80, 1995
A:Title: The primary structure of cellobiohydrazase gene (CBH II) from Trichoderma viride
A:Reference number: A38979
A:Accession: A38979
A:Molecule type: DNA
A:Residues: 1-471 <MAN>
C:Gene: cbhII
C:Genetics:
A:Introns: 31/2; 160/2; 243/1
C:Superfamily: cellulose 1,4-beta-cellobiosidase II; fungal cellulose-binding domain
C:Keywords: glycosidase, hydrolase, polysaccharide degradation
F31-62/Domain: fungal cellulose-binding domain homology <FCB>

Query Match          9.9%  Score 548;  DB 1;  Length 471;
Best Local Similarity 31.2%;  Pred. No. 2,4e-22;
Matches 156;  Conservative 60;  Mismatches 176;  Indels 108;  Gaps 17;

QY  7  SGRNRQYR-----GTRPAPSKRLRAGVLGAVSIASIVP----LMO 48
    ||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB  47  SCSITCYSDNDYISQCLPRAASSSSSTRRASTTSR-----VSPYTSKSSSAMPPEGSTTTR 101

QY  49  HPAT---AATHVNDPAGATFEVNPYMAOEVOSEANOTNATLAKMKVVSITYSTAVAMD 105
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB  102  VPPVSSGATGYSGNFVGTPWANNVYASSEVSLAIPELTGMANFAAAVAVKVSEFMVLD 161

QY  106  -----RIAANGVNGGGLTLYIDAAISQOQGTTPEVIEIVLYDLEGRCDAALAS 155
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB  162  TLDDKPLMEQGLADIRTKANKNG--NYAG-----QPVVYDLEDRDCAALAS 205

QY  156  NGELEPATAAGLOTETQYIIDPLASILSNPKYSLRIYIIIEPDSLPAVNTMNSIOACATA 215
    ||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB  206  NGEYSIADGQAKYK-NTYIDTRIQIV--VEYSDIRTLIVIEPDSLANKVNTLGPCKANA 262

```

[illegible]

```

RESULT      11
B47093
Cellulase (EC 3.2.1.4) Cend - Cellulomonas fimi
N/Alternate names: endo-1,4-beta-glucanase
C/Species: Cellulomonas fimi
C/Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 06-Dec-1996
C/Accession: B47093
J:Meinke, A.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.
J. Bacteriol. 175, 1910-1918, 1993
A:Title: Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase D (Cend),
A:Reference number: A47093; MUID:9320993
A:Accession: B47093
A:Status: Preliminary
A:Molecule type: DNA; protein
A:Residues: 1-747 <MEI>
A:Experimental source: ATCC 484
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBI:128120, NCBI:128122)
C/Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Superfamily: bacterial cellulose-binding domain homology
C:Keywords: glycosidase; hydrolysis; polysaccharide degradation
F:044744/domain: bacterial cellulose-binding domain homology <BCB>
F:045744/disulfide bonds: #status predicted

```

Query Match	8.88;	Score 487;	DB 2;	Length 747;
Best local similarity	27.18;	Pred. No. 8.9e-19;		
Matches 168;	Conservative 67;	Mismatches 175;	Indels 210;	Gaps 23;
QY	498	KRGLOLVNMGSS-----	-VDLST-----	-VTVRVPTFR 525
		: :		: :
Db	260	KRPVPTSTGILDYYGTWGNLRCGRDHPIDLGADQOLYSPHDYGLVFPQKFKQKD		319
QY	526	GGSTL--VYNDDAAMCGNIR---ASFQSVNRPATPDTYL-----	QLSFT	568
		::: :		::: :
Db	320	FDKASLADYVGNMFLFIHDEDIAPLLIGEMGRIGDOPRODKMAALRLDIAVERRLSQT		379
QY	569	GGTL-AAAGSTGELQNRVKKSDMSNTEFNNDYSGTNTTFDDMTKVTVV-----	NG	619
		: :		:
Db	380	FWVLPNPSGTGGILL--LDDMKYDEV--KYSMLPEPLTMKRGCKRYVGLDHOVPLGG		432
QY	620	V-----LYWGTEDVTPSPVPGGLVYTVGSSGVSLANASTD--NGVAHYNYR		667
		: : : :		: : : :
Db	433	VGSTGTGISQVGGGTEDTAPTAPTGLRAGPTASTVLTLSASTDGGSSVAGVEYR		492
QY	668	NCGLVAGQPTVTSFTDGLIAGTAYTYTVAAVAADGNTSAPSPVDCSTPBNQNGTVSQD		727
		: : : : :		
Db	493	GTLVGTATTATSYVTGTGLADSAVTFEVRKDKDAGNTSA-----		531

[illegible]

RESULT 12
T31337
1,4-beta-glucanase (EC 3.2.1.-) - Anaerocellum thermophilum (fragment)
C:Species: Anaerocellum thermophilum
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T31337
R:Zverlov, V.; Mahr, S.; Riedel, K.; Bronnenmeier, K.
Microbiology 144, 457-465, 1998
A:Title: Properties and gene structure of a bifunctional cellulolytic enzyme
contains.
A:Reference number: Z21003; NCID:98154434
A:Accession: T31337
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1711 <AVE>
A:Cross-references: EMBL:Z86105; NID:e1071329; PID:e350354; PIDN:CAB06786.1
C:Genetics:
A:gene: celA
C:Keywords: glycosidase; hydrolase

[illegible]


```

QY 319 D-----YAVDLYSRLLVAAAF-PSSIGMLIDTLRNGGCPNPPCP----- 357
Db 483 SGWPAKATDKLRFYVDL-SELIRAGVSPNOL-----TLSTNYOGAAGVSGPYPDASK 536
QY 358 -----STATDV-----NRFVNQSKIDLRHRL-MCNONG-----AGLGGPPQASP 397
Db 537 NIYYLLVDFTGLIRPGGODKKKEVQRIAPQVQMDNSDYFQDILKGVSSSVYKT 596
QY 398 TDFPNAHLDAVYVIRKPGESDGT-----SAASDPTTGKKSDPMCDPTT----- 441
Db 597 KYIPLYDDVKKVWGDPGSGATPPRATAPRPTVPTVPTPTPTPTPTPTPTPTPTPT 656
QY 442 -----TSYGVLTNALPNSPIAGQWPPAOFDOLVANAARAVYGGGLKQYKND 489
Db 657 TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 696
QY 490 SAPGDNQIKPGLQLVNTGSSSVLDSTVYVRYWFTRDGSSSTLVYNCDMAACGNIRASF 549
Db 697 TNSTNTIRPMUKVYNTGSSSIDLSRVTIRYWTYDGDKAQSAIS-DMAQIGASNVTFKF 755
QY 550 GSVNATPTADTYIQLSFTGT--LAAGSGTEIGNRVKNKSDMSNFTETNDYSY-GTNTT 606
Db 756 VLSSSVSGADYILEIGFKSGAGQLQAGKDTGEIQRNKSDMSNVMQNDMSMQSMTN 815
QY 607 FQDMTKRYVYVGVLYWGTEDVTPPSVPTGLVYTVGSSSVSLANNASTDNVCAHYNY 666
Db 816 YENKRYATYIDGLVWGOE-----PSGAPPT----- 842
QY 667 RAGVLVGOPTVTSFTDTGLAAGTATYTYVAADAGNTSAPSTPYDCPTGPNONGVTSYQ 726
Db 843 -----PLAT-----PAPYTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 863
QY 727 DGEYRQTNEMSSAQQCLTINTATGAMTVSTANSGTGAPATYPSIYKCHMCNCTT 786
Db 864 -----TATPTATPT----- 877
QY 787 KAVGMPRIQISQISGAVTSMTSTQVSSGANDVAY-DIWTNSTPTTGGQPGCTEIMITLNR 845
Db 878 -----SSSTPVAGQIKVLYANKETINSTNT-----TRPMLK----- 908
QY 846 GGVQPFQSGTATGTVVAGHTVWVWQOQTSMKIISVILTPGATISNLDLKAIFADAAR 905
Db 909 --VYNTGSSSIDLSRVTIRYWTYDGDKAQSAISDMAQI-GASNTV-----FKFKVLS 958
QY 906 GSLMTSDVILDYAGEFEIKMGQGLGNSFSVTSVGVACRATYVYVNSDMGSGFTATYT 965
Db 959 SSVSGADYILEI--GFKSGAGQLQAGKDTGEIQRNK-----KSDMEN----- 999
QY 966 VYNTGSRATNGTVAWSFGNCTVYNTVNTALTQSGASVYATNLVSNNTYQCGSTTTCF 1025
Db 1000 -YNOG---NDM---SMM-----OSMTNY-----GENVKVT-AVTDGLVWGOEP--- 1035
QY 1026 NGSYGTNAAPT 1037
Db 1036 -----SGAPPTPT 1043

```

```

RESULT 13
T17120
cellulase (EC 3.2.1.-) precursor, thermoactive - Caldocoelium saccharolyticum
C:Species: Caldocoelium saccharolyticum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C:Accession: T17120; A43745
R:Te'o, V.S.; Sauli, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A:title: Cella, another gene coding for a multidomain cellulase from the extreme thermoph
A:Reference number: Z18698; MUID:95336703
A:Accession: T17120
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1742 <TRQ>
A:Cross-references: EMBL:L32742; NID:g537499; PID:g537500; PIDN:AAA91086.1

```

```

R:luethl, E.; Bhana Jsmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:title: Cloning, sequence analysis, and expression in Escherichia coli of a gene cod
A:Reference number: A43745; MUID:91247819
A:Accession: A43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1516-1544, 'A', 1546-1742 <LUE>
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72860.1; PID:g144293
C:Genetics:
A:Gene: cels
C:Keywords: glycosidase; hydrolase

```

```

Query Match      8.1%; Score 449.5; DB 2; Length 1742;
Best Local Similarity 21.5%; Pred. No. 3e-16;
Matches 249; Conservative 114; Mismatches 389; Indels 407; Gaps 48;

```

```

QY 29 AGVLAGAVSIASIVPLANOHAIAATV-----DNPYAGATFFVNPY-- 71
Db 173 AVVAETASLAAASIVLKDRNPTKATYLDHAKDLYEFAEVTKRSDSGYTAANGYNSWGS 232
QY 72 -----MAQVQSEANQTNATLAAKMRVSTYSTAWMMDRTAATNGVNGGGLTYYDA 125
Db 233 FYDELSSMA-----AVWL--YLATND-----STYLFK 256
QY 126 ALSQOQGTPEVIEIYIDLPRDCALASNGE--LPATNAGIQETEQYIDPIASILSN 183
Db 257 AESYVQ--NMPKISGSNIIDYKWAHCMDVDHNGAALLAKITQDKYK-QIISHLDYMT 314
QY 184 P-----KSSLRITYIIIEPDSLPAVNTMSI-----QACATAV-----PYEQGLE 225
Db 315 GYGERIKYTPGKMLMDQMSLRATYTAFLAFYSDWSGCGTQKKETRYRFGESQIDY 374
QY 226 AL-----TKHAIPNYIYMDAHSQM-----LGMPNNSAGVYQ 259
Db 375 ALGSGRSPVFGCTNPFRPHHRTAHSMAQSOSIPSHRRTLIGALVGGGSDSTYD 434
QY 260 EVQKVLNASIGVNGIDFVYNTANTY-----TPLEPFM----- 292
Db 435 DISNYVNEVACDYNAGFALAKMYLLYGNPIPDFKAIETPTNDEFVEAGINASTN 494
QY 293 -----TATQVGGQRPESANFYQWNPDIIDEADYANDVLSRLVAAAF-PSSIGMLIDTLRN 346
Db 495 FEIKAIYVNSGWPAPRAVTKLFR-----YFVDL-SELIRAGVSPNOL-----TLST 541
QY 347 GNGGPNPEPTG-----STATDV-----NRFVNQSKIDLRHRL-MCNON 385
Db 542 NYNQGAKVSGPYVWDBSSRNIIYIINDFTGLIRPGGDKYKKEVQRIAPQVQMDNSN 601
QY 386 G-----AGLGGPPQASPDPFPNAHLDAVYVIRKPGESDGTSAAS-DPTTGKKSDPMCDP 438
Db 602 DYSFQDIKGVSSGVYKTYIPLYDEDIKW-----GEEPGTGSVSPTPASVTPPTPTPT 657
QY 439 TYTTSYGVLTNALPNSPIAGQWPPAOFDOLVANARA-----VYSGGLKQYKND 491
Db 658 TATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 717
QY 492 PGDNQIKPGLQLVNTGSSSVLDSTVYVRYWFTRDGSSSTLVYNCDMAACGNIRASFS 551
Db 718 STTNTIRPMUKVYNTGSSSIDLSRVTIRYWTYDGDKAQSAIS-DMAQIGASNVTFKF 776
QY 552 VNPATPTADTYIQLSFTGT--LAAGSGTEIGNRVKNKSDMSNFTETNDYSY-GTNTT 608
Db 777 LSSSVSGADYILEIGFKSGAGQLQAGKDTGEIQRNKSDMSNVMQNDMSMQSMTISG 836
QY 609 DMTKYVYVYVGVLYWGTEDVTPPSVPTGLVYTVGSSSVSLANNASTDNVCAHYNYRN 668
Db 837 ENEKYATYIDGLVWGOE-----PSGTTPTA----- 861
QY 669 GVLVGOPTVTSFTDTGLAAGTATYTYVAADAGNTSAPSTPYDCPTGPNONGVTSYVQD 728
Db 862 -----PTSTP-----TYTVTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 899

```


Best local similarity 22.7%; Pred. No. 2.9e-15;
Matches 211; Conservative 105; Mismatches 325; Indels 287; Gaps 40;

```
QY 44 PLAMQHPIAATHVDNPY--AGATFV--NPYMAOEVOSEANOTNAT--LAARKRVYS 96
Db 695 PTPPTPTVVSATPTPAPASAGSYWPSSEYGALKVYANGNMSSTTNVLPNKIKIEN 754
QY 97 TYSTAVMMDRIA-----AINGVNGGPGLETTYIDAA--LSQOQGTPEVTEIYIDLPGRD 149
Db 755 VGTAVVDLSRVKRYWYITDG-----EAAQSVSVASSINPAYIDVRV----- 797
QY 150 CAALASNGELPATAAGLQTY-ETQYIDPIASILSNPKYSLSRIYVTLIEPDSLPNNAVTNMS 208
Db 798 -----KLGAMAGADYVEVGFKSGAGVLAAGSTKEIR-----LS 833
QY 209 IQACATAVPEYEQIEYALTKLHAIPNYI-----YMDAHSGLGWPNN----- 253
Db 834 IOKSSGS---YQSNDSYSVRS---ANSYIENEKVYIDVY---LVMGREPGRNAQIK 882
QY 254 ---AGYVQEVQKVLNAGISGVNGIDGPTNTANTYPLKEPEFTATQVGGQPVESANFYQ 310
Db 883 VWTANGNLSSPTTNVLPNKIKIENVG--TTAVDLSRVKRYWYITDGATQSVVAS--S 937
QY 311 WNP-----DIDEADYAVDLYSLVAAGFPSSIGMLIDFLRNGMGPNNEPTGPS 358
Db 938 INPATIDRVKVLGAMAGADYVEY-----GFKSGAGVL-----AAGQS 977
QY 359 TATDVNTFVNOSKIDLRQHRGLMCNONGAGLQPPQASPTDFPNMHL--DAYVWIKPGEIS 417
Db 978 T-----KEIRLSIQKSSGSYNQSNDSYSVRSANSYIENEKVYIDVLMWKEP--S 1027
QY 418 DGTSAASDPTTGKSDPMDPDTYTSYGVLTNALPNSPIAGQFPAPQDFQLVANARPAV- 476
Db 1028 RGTKPAGEVTPAPPTSTPTPTPTTAPTSAPTPSPVTAATPTPTPTPTPTPTPTT 1087
QY 477 -----VSGGLKQYKKNDSAPGDNOIKPGLQLVNTGSSSDLS 514
Db 1088 PTPPTPTPTVTPPTPTPTGCTPGTGLKVLKNNETSASTGSIKRPFKLVNGSSSDLS 1147
QY 515 TYTVRYWFTRDGGSSTLYNCDMAAMGCGNIRASFGSVNPAATPTADTYLQLSFTGTT-L 572
Db 1148 RYKIRYWTYVDGDKPQSAV-CDMAQIGASNVPFNFKLTSVSGADYLYLEVGFSSGAGQL 1206
QY 573 AAGSGTGEIQNVKNSDMNSFTETNDYSY-GTNTFPQDMTKTYVYNGVLVWG----- 624
Db 1207 QPCKDTGDIQVFNKNDMSNYSNQADMSMQSMKTNYGENAKVTLYYDGYLVWGOEPGAT 1266
QY 625 ---TEDVTPPSVPTGIVYTVGVSGSSVSLAMNASTDNVGAHYNVYRNGVLYGQPTVTSPT 681
Db 1267 PAPPTSTATPTPTPTATVT-----PTPTA-T 1290
QY 682 DTGLAGTAYTYTVAAVDAAGNTSAPSTPVDCPTGPNONGVTSVODG----- 728
Db 1291 PTPPTPTATPTPTPT-----PTTPPTKPVGKIP-PNNNPFLSHKFGADPAVLVYGRVY 1344
QY 729 -----EYRYQVTEWSSAQOCLTINTATG---ART-----VSTANBSG 763
Db 1345 MYLTNDILEYDENGWVKDMSYSKINKITIVYSSDDLVTNWDGHEIEVAGPENGVAKMANFS 1403
QY 764 GTGGAAP-TYPSI-----YKCGHMGNCTKKNVGMPIQISGSAVTSWSTTQVSSGA 814
Db 1404 ---WAPAVAYKINKNDREFLYFG-----NSGGGIGVITADSPGTGPMWD----- 1444
QY 815 YDAVDIMTNSTPTTGTQPNGTETIMWL 842
Db 1445 -PLGRPLITWSTP-----GVQGVVWL 1464
```

Search completed: August 29, 2002, 16:24:14
Job time: 349 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM, protein - protein search, using sw model

Run on: August 29, 2002, 16:38:35 ; Search time 38.44 Seconds
(without alignments)
1050.585 Million cell updates/sec

Title: US-09-917-384-6
Perfect score: 5336
Sequence: 1 MERTQSGRCNRCTRGCTRM.....GFGSGYGTNAAPTLCSTAS 1043

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1707	30.8	872	1 GUXA_CELFI	P50401 cellulomona
2	641	11.6	438	1 GUX3_AGABI	P49075 agarius bl
3	586	10.6	1090	1 GUXB_CELFI	P50899 cellulomona
4	561.5	10.1	1045	1 GUXB_CELFI	P26225 cellulomona
5	561	10.1	462	1 GUXB_FUSOX	P46236 fusarium ox
6	548	9.9	471	1 GUX2_TIRRE	P07987 trichoderma
7	487	8.8	747	1 GUXD_CELFI	P50400 cellulomona
8	449.5	8.1	1742	1 GUNB_CALSA	P22534 caldocellum
9	434	7.8	1331	1 GUNB_CALSA	P22533 caldocellum
10	414	7.5	562	1 GUN1_ACICE	P25833 actinobacteriu
11	413	7.5	1853	1 GUN1_ACICE	P06851 clostridium
12	401	7.2	772	1 GUNB_CLOCT	P01866 clostridium
13	388.5	6.9	1039	1 GUNB_CALSA	P29719 paenibacilli
14	383.5	6.9	700	1 GUNB_CALSA	P10474 c endoglucan
15	382	6.9	499	1 GUN3_BACSU	P10475 bacillus su
16	381.5	6.9	499	1 GUN3_BACSU	P23549 bacillus su
17	380	6.9	145	1 YCEA_PAEIA	P29718 paenibacilli
18	379	6.8	499	1 GUN1_BACSU	P07983 bacillus su
19	366.5	6.6	880	1 GUN1_THERFU	P26221 thermomonas
20	363.5	6.6	504	1 GUN1_THERFU	P26222 thermomonas
21	355.5	6.4	505	1 GUN1_THERFU	P26222 thermomonas
22	338.5	6.1	441	1 GUN2_CLOS	P26539 clostridium
23	325.5	5.9	986	1 GUN2_CLOS	P20533 bacillus cl
24	315	5.7	699	1 CHIA_BACCI	P38058 clostridium
25	309	5.6	1848	1 GUN2_CLOCT	P50900 clostridium
26	307.5	5.6	914	1 GUN2_CLOCT	P26414 microspor
27	299.5	5.4	456	1 GUN1_MICBI	P27050 bacillus cl
28	289	5.2	879	1 GUN1_CLOCT	P02934 clostridium
29	276.5	5.0	524	1 GUN1_CLOCT	P27050 bacillus cl
30	273.5	4.9	962	1 GUN1_PSEPL	P10476 pseudomonas
31	261	4.7	488	1 PHB_ALCPA	P12625 alcaligenes
32	257.5	4.6	2021	1 OMPA_RICCN	P02657 rickettsia
33	253	4.6	484	1 GUX_CELFI	P07986 cellulomona

34	248.5	4.5	449	1 GUN1_CELFI	P07984 cellulomona
35	244	4.4	644	1 XYND_CELFI	P54855 cellulomona
36	235.5	4.3	1148	1 ICER_PSEEX	P03611 pseudomonas
37	234	4.2	237	1 GUN1_ASPEC	P22669 aspergillus
38	234	4.2	597	1 CHIA_STROI	P05638 streptomyc
39	233.5	4.2	1200	1 ICEN_PSEEX	P06620 pseudomonas
40	230	4.2	619	1 GUN1_PSEEX	P36909 streptomyc
41	228.5	4.1	748	1 GUN1_PSEEX	P27033 pseudomonas
42	228.5	4.1	1196	1 ICEN_PSEEX	P03479 pseudomonas
43	226.5	4.1	1861	1 APU_THERFU	P38336 t amylopull
44	226	4.1	610	1 CHIT_STRPL	P11220 streptomyc
45	224.5	4.1	515	1 GUN1_CLOCT	P28623 clostridium

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	872 AA.
GUXA_CELFI				
ID	GUXA_CELFI			
AC	P50401:			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Exoglucanase A precursor (EC 3.2.1.91) (Exocellulohydrolase A)			
DE	(1,4-beta-cellulohydrolase A) (CBP95).			
GN	CBHA.			
OS	Cellulomonas flm1.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.			
OX	NCHI_TaxID=1708;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 484;			
RX	MEDLINE=94344030; PubMed=8065260;			
RA	Medlne A., Gilkes N.R., Kwan E., Kilburn D.G., Warren R.A.J.,			
RA	Miller R.C., Jr.;			
RT	"Cellulohydrolase A (Cbha) from the cellulolytic bacterium			
RT	Cellulomonas flm1 is a beta-1,4-exocellulohydrolase analogous to			
RT	Trichoderma reesei CBH II."			
RL	Mol. Microbiol. 12:413-422(1994).			
RN	(2)			
RP	SEQUENCE OF 41-58.			
RX	MEDLINE=93209933; PubMed=8458833;			
RA	Medlne A., Gilkes N.R., Kilburn D.G., Miller R.C., Jr., Warren R.A.J.;			
RT	"Cellulose-binding polypeptides from Cellulomonas flm1: endoglucanase			
RT	D (Cend), a family A beta-1,4-glucanase."			
RL	J. Bacteriol. 175:1910-1918(1993).			
CC	-1- FUNCTION: THIS ENZYME HYDROLYSES 1,4-BETA-D-GLUCOSIDIC LINKAGES OF			
CC	CELLULOSE. WEAK ACTIVITY AGAINST CARBOXYMETHYLCELLULOSE, BACTERIAL			
CC	MICROGELATINASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS WITH INVERSION			
CC	OF ANOMERIC CONFIGURATION.			
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages			
CC	in cellulose and cellotetraose, releasing cellobiose from the non-			
CC	reducing ends of the chains.			
CC	-1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN			
CC	(CBD).			
CC	-1- SIMILARITY: BELONGS TO CELLULASE FAMILY B (FAMILY 6 OF GLYCOSYL			
CC	HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: L25809; AAC36898.1; -			
DR	HSSP: P07986; IEXG.			
DR	Interp: IP001919; CBD_2.			

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

	EMBL:	L24519;	AAA50607.1;	-
	EMBL:	L24520;	AAA50608.1;	-
	EMBL:	Z34007;	CNA83971.1;	-
	HSSP:	P00725;	IAX6.	
	InterPro:	IPR00254;	cbd_fungal.	
	InterPro:	IPR01524;	Glyco_hydro-6.	
	Pfam:	PF00734;	cbd_1; 1.	
	Pfam:	PF01341;	Glyco_hydro-6; 1.	
	PRINTS:	PR00733;	GLHYDRASE6.	
	PRODOM:	PD001821;	cbd_fungal; 1.	
	PRODOM:	PD003733;	Glyco_hydro-6; 1.	
	SMART:	SM00236;	fcbd; 1.	
	PROSITE:	PS00562;	cbd_FUNGAL; 1.	
	PROSITE:	PS00655;	GLYCOSYL_HYDROL_F6_1; 1.	
	PROSITE:	PS00656;	GLYCOSYL_HYDROL_F6_2; 1.	
	Cellulose degradation:	Hydrolase;	Glycosidase; Signal.	
KW	SIGNAL	1	20	POTENTIAL.
FT	CHAIN	21	438	EXOGLUCANASE 3.
FT	DOMAIN	21	59	CELLULOSE-BINDING (BY SIMILARITY).
FT	DOMAIN	60	87	LINKER.
FT	DOMAIN	88	438	CATALYTIC.
FT	ACT_SITE	215	215	PHOTON DONOR (BY SIMILARITY).
FT	ACT_SITE	393	393	NITROPHILE (BY SIMILARITY).
FT	ACT_SITE	28	45	BY SIMILARITY.
FT	DISULFID	39	39	BY SIMILARITY.
FT	DISULFID	170	229	BY SIMILARITY.
FT	DISULFID	360	407	BY SIMILARITY.
FT	DISULFID	133	133	V -> T.
FT	VARIANT	152	152	R -> Q.
FT	VARIANT	244	244	V -> I.
FT	VARIANT	248	248	N -> D.
FT	VARIANT	398	398	N -> S.
SEQUENCE		438 AA:	46209 MM:	002C9J3544893794 CRC64;

[illegible]

Dd 352 IRDWMGDMCNVAGAAFGVAAT 280

OY 434 PMCDPTTTSYGVLNALNPSBIAGOMFPAQFDOLVANAPAV 476
 | : : : | | | | | : | | | | :
 | : : : | | | | | : | | | | :

Dd 405 SHCS-----LSDAHPPAPEACTWTFQAFFETLVANANPAL 438

```

RESULT 3
ID      GUXB_CELFI STANDARD:      PRT: 1090 AA.
AC      P50899:
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      Exoglucanase B precursor (EC 3.2.1.91) (Exocellulohydrolase B)
DE      (1,4-Beta-cellulohydrolase B) (CBP120).
DE      CNBH OR CENE.
OS      Cellulomonas flm1.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
OX      NCBI_TaxID=1708:
RP      [1]
RP      SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.
RC      STRAIN=ATCC 484;
RC      MEDLINE=96003898; Pubmed=7575482;
RA      Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT      "Cellulohydrolase B, a second exo-cellulohydrolase from the
RT      cellulolytic bacterium Cellulomonas flm1.";
RL      Biochem. J. 311:67-74(1995).
RN      [2]
RN      SEQUENCE OF 54-75.
RP      MEDLINE=93209933; Pubmed=8458833;
RA      Melnik A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT      "Cellulose-binding polypeptides from Cellulomonas flm1: endoglucanase
RT      D (Cend), a family A beta-1,4-glucanase.";
RL      J. Bacteriol. 175:1910-1918(1993).
RN      [3]
RN      SEQUENCE OF 54-78.
RP      MEDLINE=94197708; Pubmed=8147863;
RA      Shen H., Tomme P., Melnik A., Gilkes N.R., Kilburn D.G.,
RT      Warren R.A.J., Miller R.C. Jr.;
RT      "Stereochemical course of hydrolysis catalysed by Cellulomonas flm1
RT      Cend, a member of a new family of beta-1,4-glucanases.";
RL      Biochem. Biophys. Res. Commun. 199:1223-1228(1994).
CC      -1- FUNCTION: HYDROLYSE CELLULOSE TO A MIXTURE OF CELLOTREROSE,
CC      CELLOTRIOSE AND CELLOBIOSE, WITH ONLY A TRACE OF GLUCOSE. IT
CC      HYDROLYSE CELLOTREROSE TO CELLOBIOSE, BUT IT DID NOT HYDROLYSE CELLOTRIOSE.
CC      CELLOTREROSE AND CELLOBIOSE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS
CC      WITH INVERSION OF ANOMERIC CONFIGURATION.
CC      HAS ALSO WEAK ENDOLUCANASE ACTIVITY.
CC      -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC      in cellulose and cellobiose, releasing cellobiose from the non-
CC      reducing ends of the chains.
CC      -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC      -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC      (CBD).
CC      -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY L (FAMILY 48 OF GLYCOSYL
CC      HYDROLASES).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: L38827; AAB00822.1; -
DR      HSSP: P07986; IEXG.
DR      InterPro: IPR001919; CBD_2.
DR      InterPro: IPR003961; FN_III.
DR      InterPro: IPR003962; FNIII_repeat.
DR      InterPro: IPR000556; Glyco_hydro_48.

```

DR Pfam: PF00553: CBD_2; 1.
 DR Pfam: PF00041: fn3; 3.
 DR Pfam: PF02011: Glyco_hydro_48; 1.
 DR PRINTS: PR00014; ENTRYELIT.
 DR PRINTS: PR00844; GLHYDRLASE48.
 DR PRODOM: PD011903; Glyco_hydro_48; 1.
 DR SMART: SM00060; FN3; 3.
 DR PROSITE: PS00561: CBD_BACTERIAL; 1.
 DR Cellulose degradation; hydrolase; Repeat; Signal.
 KM SIGNAL
 FT PROPEP 1 33
 FT CHAIN 34 53
 FT CHAIN 54 1090
 FT DOMAIN 54 699
 FT DOMAIN 700 785
 FT DOMAIN 794 884
 FT DOMAIN 891 978
 FT DOMAIN 989 1090
 FT ACT_SITE 513 513
 FT DISULFID 990 1089
 SQ SEQUENCE 1090 AA; 114829 MW; 046BB9D562F399 CRC64;

Query Match 10.6%; Score 586; DB 1; Length 1090;
 Best Local Similarity 24.0%; Pred. No. 4.3e-25;
 Matches 299; Conservative 147; Mismatches 419; Indels 382; Gaps 62;

QY 20 MAIAKRLAGYAGAVSIASTVPLAMOHPAIAATHDNPAGATFFVNPYMAOEVOSE 79
 DB 1 MSSTRRRSAMVAATVGV-SSFLAVAGITPAIAAGAGQP---AT-----VWVP 46
 QY 80 ANQTNATLAAK--MRVSYSTAVMMORIAINGVSGPLTTLDAALSOOGGTPPEV 137
 DB 47 AASPRAAVDGEYAOEFLAQY-----DKIK--DPANG-----YFSA-----QGLPPIHA 87
 QY 138 IEIVTYDLP--GRDCALASN-----GELPATAAGL-QTYET-----QYIDPIA 178
 DB 88 VETLAEAPDYGHETTSEASYWMLLEALYGOVQDMAPLNMAMPTMEKYMIPQSVDPQT 147
 QY 179 SILSNPKYSSLRIVITIEPDSLPAVNTN-----MSIQACATAPVYEGGIEFALTKL 230
 DB 148 NSFYFNPSPATYAPPEFNHSSYSPSOINSGISGTDPIGELKAT-----YGNADVOMHWL 203
 QY 231 HAIPVYIYMDAAHSG-WLGMFNNAAGYQVQKVLNLSI-----GVNGCI 274
 DB 204 AAVDNIYTGATPGACCTIGPATGTSPINTPQRBQESVWETVPOPSCEEFKYGSKCY 263
 QY 275 DGFVTNTATYPLKEPFWATQVGGQVPS---ANFYQWNPDIIEA----- 318
 DB 264 LDLETKDASYA--KQWKYTSASDADARAFAEAYWMAN--QWATEQGAADVAATVAKAKM 319
 QY 319 -Y-----AVDLYSRVY-----AAGPSSIGMLDITLNGNGGPRE----- 353
 DB 320 GYLKRYTLEFDKYFKKIGCTSPCAAGQGRHAHYLLSWYMAAGCATDSSGNAMRIGSSH 379
 QY 354 -----PTGPSTATDVNTFVNQSKIDLRQHRGLCMONGAGL--- 389
 DB 380 AHFGYONPLAAMALSTDRLTKPSPTAKADMAAM-QRLLEY-----TWLQASNGCIAG 434
 QY 390 -----GQPPQASPTDFPNAHLDAVYWKIPGES-DGTSA-----ASDPTT 428
 DB 435 ATNSMDGAVAGPAGTPTFYGMGYTEAPYVDPPSRMRGMQAMGQVRAVELYASGNMQ 494
 QY 429 GKSDPMCDPTTYSYGVLTNALPNSPIAGQMPPAQFDOLVAMNAPVYSGGLKYQYK- 487
 DB 495 AAK-----IIDKWPVWVAVNISTDGASWKPSEIKTKWGKPD 530
 QY 488 --NDASPGDNQIKGTLQ-LVNTGSSVYDLSVTYVR---YWFTRDGGSSST---LYNYND 536
 DB 531 TWMAAAPFTN---PGLITVEVTSYGQDVGYAADTARALLFYAAKSGDTARAKALLDI 587
 QY 537 WAMAGCGNIRASGVSVPATPTADTYTLQSLFT-----GGTLAAGSGTEGION-RVKKSDM 590
 DB 588 WA-----NNQDPLGV--SAVETRGDYKKRPDDTYVANGSDIYIPSGWTGIMPNGDVYIKPEV 640

QY 591 SNETETNDYSTGTNTTFQDMTKVTVYNG-----VLYWG----- 624
 DB 641 SFLDIRSEFYKKDPN-----WSKVQTFELDGAEPQFRYHRFMAQTVAVAGALADYARLPDQ 695
 QY 625 --TEDVTPSPVPTGLVTVYNGSSSVSLAMNASTDVGVAHVYVYNGVLYGQPTPTSTPD 682
 DB 656 TTPPTPTAPVPTGLQAGVYVSTENATISWTASTDTRVYTGIDVYKATVGTATTSTFD 755
 QY 683 TGLAAGTAVTVYVAADVADAGNTSAPS-----TPVDCPTGPNQNGVTSVQDEYHVQ 733
 DB 756 TGLTASTAVYAVYVRAFDAGVNSAPSALVTYTKATPSD-TTAPSVAPITS----- 805
 QY 734 TNEHNSAQOCLITITAGATVSTANSSGGTGGAPATYPSLYKCCMHGNCCTTKVGMPT 793
 DB 806 -----SSS-----TANSVITGMSASTDN-AGGSG-LAGY-DVYRGA-----T 839
 QY 794 QISQISAVTSWSTQVSSGAV-----DVAYDI-----WNSTPTTGQ--P 833
 DB 840 RVAQT-TALITFDITGLTASTAVETVRADVAGVNSAPSTAVSTTKSDTPTDITAPSPV 898
 QY 834 NCTEIMIMLSRGVOPRGSQATGTVVAGHTVNVMOGQTSWKIISVYLPAGATSISNL 893
 DB 899 AGLAAM-----TVTETSYA-LWNV-----ASTDTGGS 924
 QY 894 DIKA--IFRADAARSLNTSYL--LDVEAGFE-----IWGGGGLGNSNSVSYSVS--- 941
 DB 925 GLKGYDVYRGATRVGTSYASTYTDGLTAATAYQYTVRATDAGVNSAASALSYTKTP 984
 QY 942 -GVVACRAITYYVNSWMSGFTATVTVNTGSRATGVMVMSFGNGQVYNTNMTNALQS 1000
 DB 985 QTGSGCSVAIYANSS-WNSGFTASVRIINTGITTINGMSIGFLTGQKRVQOGMSATWTQS 1043
 QY 1001 GASVTATNLYSNNVIOPOGSTTFGFNGSYSGTNAAPT-----LSCT 1041
 DB 1044 GSTVATATNAPMNGTLAPGQTVDFGNGSHSGNPNPASPFTLNGACCT 1090

RESULT 4
 GUNB_CELFI STANDARD; PRT; 1045 AA.
 AC P26225;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase B)
 GN (Cellulase B).
 OS Cellulomonas fimi.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococciaceae; Cellulomonadaceae; Cellulomonas.
 OX NCBI_TaxID=1708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91100298; PubMed=1987122;
 RA Melinke A., Braun C., Gilkes N.R., Kilburn D.G., Miller R.C. Jr.,
 Warren R.A.J.;
 RT "Unusual sequence organization in CenB, an inverting endoglucanase
 from Cellulomonas fimi.";
 RT J. Bacteriol. 173:308-314(1991).
 RN [2]
 RP MEDLINE=92041609; PubMed=1938913;
 RA Melinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
 RT "Multiple domains in endoglucanase B (CenB) from Cellulomonas fimi:
 functions and relatedness to domains in other polypeptides.";
 RT J. Bacteriol. 173:7126-7135(1991).
 CC -FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGUCCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOIOHYDROLASES THAT CUT THE DISACCHARIDE CELLULOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER

SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
 CC CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- MISCELLANEOUS: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A
 CC POTENTIAL SITE FOR PROTEOLYSIS.
 CC -1- MISCELLANEOUS: MAY CONTAIN A SECOND CYPE IN THE CATALYTIC DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC -----
 CC EMBL: M64644; AAA23086.1; -
 CC PIR: A39199; A39199.
 CC HSSP: P26221; ITP4.
 CC InterPro: IPR001919; CBD_2.
 CC InterPro: IPR001956; CBD_3.
 CC InterPro: IPR003961; FN_III.
 CC InterPro: IPR001701; Glyco_hydro_9.
 CC Pfam: PF00553; CBD_2; 1.
 CC Pfam: PF00542; CBD_3; 1.
 CC Pfam: PF00541; fn3; 3.
 CC Pfam: PF00759; Glyco_hydro_9; 1.
 CC SMART: SM00060; FN3; 3.
 CC PROSITE: PS00561; CBD_BACTERIAL; 1.
 CC PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
 CC PROSITE: PS00592; GLYCOSYL_HYDROL_F9_2; 1.
 CC PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
 CC Cellulose degradation: Hydrolase; Glycosidase; Repeat; Signal.
 CC POTENTIAL.
 CC KMW
 CC SIGNAL 1 33
 CC FT SGAIN 34 1045
 CC FT CHAIN 34 643
 CC FT DOMAIN 644 650
 CC FT DOMAIN 651 733
 CC FT DOMAIN 734 748
 CC FT DOMAIN 749 830
 CC FT DOMAIN 831 846
 CC FT DOMAIN 847 930
 CC FT DOMAIN 931 944
 CC FT DOMAIN 945 1045
 CC FT ACT_SITE 410 410
 CC FT ACT_SITE 449 449
 CC FT ACT_SITE 458 458
 CC FT DISULFID 946 1044
 CC FT SEQUENCE 1045 AA; 108990 MW; AC2F7B84E4E3C4F0 CRC64;
 SQ
 Query Match 10.1%; Score 561.5; DB 1; Length 1045;
 Best Local Similarity 26.7%; Pred. No. 9,3e-24;
 Matches 265; Conservative 105; Mismatches 382; Indels 241; Gaps 47;

Db 390 DNPSSSYVGE-----GANPPTAPHHRTAHGSMWLDSTTPPAQSRHVL 433
 QY 382 CNONGAGLCQPPQASPTDFPNANHLA---YWIKPPGESDSTSAASPTTGKSDKPCMP 438
 Db 434 ---GALVGGP---GSPND---AYTDSRODY-----ANVATDYNAG----- 466
 QY 439 TTYTSGVLTNLNLPSPINGWEPNO-----FDOLVANARPAVSGGLKQYKNNDSAP 492
 Db 467 -FTSALARLVEEGCGRPFLPPEQPDGDLFEAMLNQPSGTFTEVKMIRNQSAPF 525
 QY 493 GDNQIKPGLQLVGTSSVDLSVYRYWFRDG--GSITVYNCDMAAGCGNIRASFGS 551
 Db 526 ARS-----LNNAKRYFTTDFGFAASVTLNAYSE--CG--AOSGK 563
 QY 552 VNPATPADYTLQLSFTGTLAAGST---GEIQNRV--KKSQMSNFTEDNOVSYG--TNT 605
 Db 564 GVSAGGTIG--YVELSCVGODIHFGGSOHRETOFLTGDPAGN---PANDSYGLQYT 619
 QY 606 TFDQMTKVTYVNGVLYWG-----TEDVTPSPVPTGLVYTVGSSSVLANNASTD--NV 658
 Db 620 ALAKASATITLVGSLVWGKEPPTGTTDTTPPTTGPATGVTGASISMAASTDACS 679
 QY 659 GVAHNVYR---NGVLYGQPTVTSFTDTGLAGTATTVVAVDAAAGNTASPTPVDCT 714
 Db 680 GVAGIELRVGCTTQTLVGTTLAAAYILRLDLPATASVYKAKDAVAGNVSASAVT-- 737
 QY 715 PCPNQGVTSVODGRTYRQTNEMNSAQOCLITWATGAMTVSTANPSCGTGAPATPDS 774
 Db 738 -----FTTDTTGTETPEPTTP--GTPVASAVTSGATLAMPST-----GDPAN--S 779
 QY 775 IYKCGHMGNCITKKNVGMPIQISQISAVTSWSTQVSSGADVADYDINTSTP---TTTG 831
 Db 780 GYVVLAVQCTTTTVAQAQTVPTVLSGLTPSAVYAVRAKNVAGDVSAVPTTTAA 839
 QY 832 QPNGTEIMTLMSKRGVOVF--GSQTATGYVYAGTMMVMOGQOTSW--KIISVILPGA 887
 Db 840 PVDVT-----VAFTVPTPYASNAVATGAIL--TWNASTDSGGSGLAGLEVLRVS 887
 QY 888 TSISNLDLKAIFDAARSLNTSDVLDLVEAGFEIWOGGCGIGNSFSVSTGGV--- 944
 Db 888 GTTQTLVASTTMTVALAGLTPATASVYVRAK---DAGNVSASVSPVFTLLPYST 943
 QY 945 -ACRATTVNSDMGSGFTATVYTNATNGVATNGVAFSGNGQVYTNMTALTQSGAS 1003
 Db 944 PSCFTVAVSTNS--NMGVGTSGVKITNTGTTPLT--WTIGFAPSGQOVYQGSATWSQTGT 1001
 QY 1004 VTATNTVSNVYIOPGQSTFEFGNSYSTNAAP 1036
 Db 1002 VTATGISMNATLQPGQSTDIGFNGSHPTNTNP 1034
 RESULT 5
 GUNB_FUSOX STANDARD: PRT: 462 AA.
 AC P46236:
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Putative endoglucanase type B precursor (EC 3.2.1.4) (Endo-1,4-beta-
 DE glucanase) (Cellulase).
 OS Fusarium oxysporum.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; mitiosporic Hypocreales; Fusarium.
 OC NCBI_TaxID=5507;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95047531; PubMed=7959045;
 RA Sheppard P.O., Grant F.J., Oort P.J., Sprecher C.A., Foster D.C.,
 RA Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.;
 RT "The use of conserved cellulase family-specific sequences to clone
 RT cellulase homologue cDNAs from Fusarium oxysporum.";
 RL Gene 150:163-167(1994).


```

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: CONTRAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL: I29377; AAA65585.1; -.
DR HSSP: P07987; ICH2.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR001524; Glyco_hydro_6.
DR Pfam: PF00734; CBD.1.1.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PR00733; GLYDRLASE6.
DR ProDom: PD001821; CBD_fungal; 1.
DR ProDom: PD003733; Glyco_hydro_6; 1.
DR SMART: SM00236; ICBID; 1.
DR PROSITE: PS00562; CBD_FUNGAL; 1.
DR PROSITE: PS00555; GLYCOSYL_HYDROL_F6_2; 1.
DR PROSITE: PS00556; GLYCOSYL_HYDROL_F6_2; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
KW SIGNAL
FT CHAIN 1 16
FT DOMAIN 17 462
FT DOMAIN 17 65
FT DOMAIN 66 99
FT DOMAIN 100 462
FT ACT_SITE 190 190
FT ACT_SITE 236 236
FT ACT_SITE 416 416
FT DISULFID 33 50
FT DISULFID 44 60
FT DISULFID 191 250
FT DISULFID 383 430
FT CARBOHYD 37 37
FT CARBOHYD 223 223
FT CARBOHYD 272 272
FT CARBOHYD 317 317
SQ SEQUENCE 462 AA; 49207 MW; E25B2F5B828B637F CRC64;

Query Match 10.1%; Score 561; DB 1; Length 462;
Best Local Similarity 32.5%; Pred. No. 3.4e-24;
Matches 150; Conservative 71; Mismatches 167; Indels 74; Gaps 16;

17 TTRMPAISRRLRAGVLAGAVSIASTIVPIAMQHPATAATHVDPYAGATFFVNPVWAOEV 76
17 TAAGSSTATATKTTATGSGSTTGGSVTSA--PPAAS--DNPYAGVDLMANNYRSREV 123
77 GSEAA---NQTNATLAAKKRVSTYSTAVMDRIAMINVGNGPGLTTLIDALSSQOQT 133
124 MNLAVPKLSGAKATAAKADVPSEF--WMDYDHISLME---DTLADIRKANKAG- 174
134 TREVEIIVYDLPGRDCALASNGELPATRAGIQETETGYIDPIASILSNPKYSRLIYV 193
175 GYVAGGFVYDLPNRDCAAAANGESLDKDKANKYKA-YIAKIKIILN--YSDTKVIL 231
194 IIEPDSLPAVNTMSIQACATVAPYEEGIEVALKRLHAIPVYIYMDAAHSGMLGMPNN 253
232 VIEPDSLAVNTLVNLDKCAKAEASAKELTYAIKELN-LPNVSMYLDAGHGMLGMPAN 290
254 ASGYVQEVQVNLASGVNGIDGFTVNTANTYPLKEPFMTATQOVGGQVESANFYQAMP 313
291 IEPAAVLVQIYKKDAKPRVRGLVTVNSNYN-----GWLKSTRPDYTESNP 337
314 DIDEADYAVDLISRLVAAAGPSSIGMLIDTLRNGMGNGPEPTSPATVNFVNGSKID 373

```

```

DB 338 NYDEORY-INAPAPLAEQSGSNVAKRIVDQGRSG-----KQPIG----- 375
QY 374 LNHQRLGMCNQGAGLGQPPQASPTDPPNAHLDAVYMIKRPESDQTSASPPTGKSD 433
DB 376 -KQAGDMCNKAGTGGFLRPSTWTGD--ALDAFVWVAPGGSDDOTS-----DTSARVD 427
QY 434 PMDDPYTTSYGLVTALNPSPIAGWEPFAQDQDVANRPA 475
DB 428 YHCG-----LDDALKPAPAGTGWFOAYFEQLDNNMPS 460

RESULT 6
GXU2_TIRE STANDARD; PRI; 471 AA.
AC P07987;
DR 01-AUG-1988 (Rel. 08, Created)
DR 01-AUG-1988 (Rel. 08, Last sequence update)
DR 01-MAR-2002 (Rel. 41, Last annotation update)
DE Exoglucanase II precursor (EC 3.2.1.91) (Exocellulohydrolase II)
DE (CBHI) (1,4-beta-cellulohydrolase).
GN CBH2.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VTT-D-80133;
RX MEDLINE=87248061; PubMed=3596237;
RA Teeri T.T., Lehtoavara P., Kauppinen S., Salovuori I., Knowles J.;
RT "Homologous domains in Trichoderma reesei cellulytic enzymes: gene
RT sequence and expression of cellulohydrolase II.";
RL Gene 51:43-52(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=QM9414 / RUT C-30;
RX Chen C.M., Gritzali M., Stafford D.W.;
RT "Nucleotide sequence and deduced primary structure of
RT cellulohydrolase II from Trichoderma reesei.";
RL Biotechnology 5:274-278(1987).
RN [3]
RP SEQUENCE OF 25-44.
RA Paegerstam L.G., Petersson L.G.;
RT "The 1,4-beta-glucan cellulohydrolases of Trichoderma reesei QM
RT 9414.";
RL FEBS Lett. 119:97-100(1980).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=90333255; PubMed=2377893;
RA Rouvinen J., Bergfors T., Teeri T.T., Knowles J.K.C., Jones T.A.;
RT "Three-dimensional structure of cellulohydrolase II from Trichoderma
RT reesei.";
RL Science 249:380-386(1990).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=97029636; PubMed=8873646;
RA Koivula A., Reinikainen T., Ruohonen L., Valkeajarvi A.,
RA Claessens M., Telemann O., Kleywegt G.J., Sardenings M., Rouvinen J.,
RA Jones T.A., Teeri T.T.;
RT "The active site of Trichoderma reesei cellulohydrolase II: the role
RT of tyrosine 169.";
RL Protein Eng. 9:691-699(1996).
CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOCELLULOHYDROLASES THAT CUT THE DISACCHARIDE CELLULOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellooligose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -1- SUBCELLULAR LOCATION: Secreted.

```


CC -1- MISCELLANEOUS: J. RESEI PRODUCES TWO DIFFERENT
 CC EXOCELLULOBIODHROLASES. THEY ARE UNIQUE IN THAT THEY CAN HYDROLYSE
 CC CRYSTALLINE CELLULOSE IN THE ABSENCE OF ENDOGLUCANASES.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
 CC HYDROLASES).
 CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.slb.ch).
 CC
 CC EMBL: M16190; AAA34210.1; -
 CC EMBL: M55080; AAA72922.1; -
 CC PIR: A26472; A26472;
 CC PIR: A26160; A26160;
 CC PDB: 3CBH; 15-JAN-91.
 CC PDB: 1CB2; 19-MAR-99.
 CC InterPro: IPR000254; CBD_fungal.
 CC InterPro: IPR001524; Glyco_hydro-6.
 CC Pfam: PF00734; CBD_1; 1.
 CC Pfam: PF01341; Glyco_hydro-6; 1.
 CC PRINTS: PR00733; GHYDRASE6.
 CC PRODOM: PD001821; CBD_fungal; 1.
 CC PRODOM: PD003733; Glyco_hydro-6; 1.
 CC SMART: SM00236; ICB0; 1.
 CC PROSITE: PS00562; CBD_FUNGAL_F6_1; 1.
 CC PROSITE: PS00655; GLYCOSYL_HYDROL_F6_2; 1.
 CC PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
 CC Cellulose degradation: Hydrolase; Glycosidase; Glycoprotein; Signal;
 CC 3D-structure. 1 24
 CC SIGNAL. 1 24
 CC CHAIN. 1 24
 CC DOMAIN. 25 471
 CC DOMAIN. 66 106
 CC DOMAIN. 107 471
 CC ACT_SITE. 199 199
 CC ACT_SITE. 245 245
 CC ACT_SITE. 425 425
 CC ACT_SITE. 425 425
 CC DISULFID. 34 51
 CC DISULFID. 45 61
 CC DISULFID. 200 259
 CC DISULFID. 392 439
 CC CARBOHYD. 38 38
 CC CARBOHYD. 111 111
 CC CARBOHYD. 121 121
 CC CARBOHYD. 130 130
 CC CARBOHYD. 133 133
 CC CARBOHYD. 134 134
 CC CARBOHYD. 139 139
 CC CARBOHYD. 313 313
 CC CARBOHYD. 334 334
 CC CARBOHYD. 199 199
 CC MUTAGEN. 245 245
 CC MUTAGEN. 359 359
 CC CONFLICT. 449 449
 CC CONFLICT. 449 449
 CC SEQUENCE. 471 AA; 49653 MW; CA711BC35B1BD88 CRC64;
 SO
 Query Match 9.9%; Score 548; DB 1; Length 471;
 Best Local Similarity 31.2%; Pred. No. 1.8e-23;
 Matches 156; Conservative 60; Mismatches 176; Indels 108; Gaps 17;
 QY 7 SGARCRQR-----GTRMPAISKRLNAGVLAGAVSTIASTIVP-----LAMQ 48
 DB 47 SGTCTVSNNDYSGCLPGAASSSSSTRRASTTR-----VSPTSRSSSATPPPGSTYTR 101
 QY 49 HPAI---AATHVDPYAGATFFVNPYAOQVSEAAQNTAATAAKRYSTYSTAVMD 105
 DB 102 VPPVSGGATATSGNPFVGVTPWMAAVYASVSSIAIPSLTGAMATMAAAVAAYKPSFMWD 161

QY 106 -----RTAALNGVNGGRLTYLDAALSOOGCTTPEVIEIYIYDLPGRDCAALAS 155
 DB 162 TLDDKPLMBQTLADIRTKARKNG--NTRG-----GFVYVDLPDRCCALAS 205
 QY 156 NGEPLPMTAGLQYETQYIDPIASILSNPKYSLRTVITTEPDSLPNAVTKMSIQACATA 215
 DB 206 NGEYSIADGVAKYK-NYIDTIRQIV--VEYSDIRTLVIEPDSILANLVTLGPKKANA 262
 QY 216 VPEYEGIEYALTKLAIRNYIYMDAHSGLMGPNNASGYVEQVQVYLNASIGVNGID 275
 DB 263 QSAVLEICINAYVQLN-LPNAVYTLDRGAGHGLGWPANODPAQLFANVYKNASSPRALR 321
 QY 276 GFVNTANYPYLPKEPMTATQVGGQPVESANFYQNNPIDADYAVDLXSLVAAGPPS 335
 DB 322 GLATNANYN-----GMNITSPSYTQGNAYYNEKLYTHALGPLLANHGW-S 367
 QY 336 SIGMLIDTLRNGMGCPNEPPTATDVNFVNGSKIDLRHGLMCNONGALGQPPQA 395
 DB 368 NAFPTDQGRSG---KQPTG-----QQQMGDMCNVIGTGFIRPSA 405
 QY 396 SPTEPPNHLDAVYKIRPGESDGTSAASDPTTGKSDKPCDPTTYSYGLNALPNSP 455
 DB 406 NTGD---SLDSFVWVVRKGGEGDGTSDSSAP---RFDShC-----ALPDALQAPAP 449
 QY 456 IAGQWPPAQFDQVYANARPA 475
 DB 450 QAGAWFOAYFYOLLTNAPS 469
 RESULT 7
 GUND.CELFT STANDARD. PRT: 747 AA.
 AC GUND.CELFT
 AC P50400;
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 01-OCT-1996 (rel. 34, Last annotation update)
 DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase).
 GN CEND.
 OS Cellulomonas faml.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
 OX NCBI_TaxID=1708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=93209933; PubMed=8458833;
 RA Weinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.
 "Cellulose-binding polypeptides from Cellulomonas faml.: endoglucanase
 D (Cend), a family A beta-1,4-glucanase."
 J. Bacteriol. 175:1910-1918(1993).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- PATHWAY: CELLULOSE DEGRADATION.
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
 CC (CBD).
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.slb.ch).
 CC
 CC EMBL: L02544; AAA3089.1; -
 CC HSSP: P07986; IEXG.
 CC InterPro: IPR001919; CBD_2.
 CC InterPro: IPR003961; FN_III.
 CC InterPro: IPR003962; FNIII_repeat.

InterPro: IPR001547; Glyco_hydro_F5.
 DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF00150; cellulase; 1.
 DR Pfam: PF00041; fn3; 2.
 DR PRINTS: PR00014; ENTPYELII.
 DR SMART: SM00060; FN3; 2.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
 FT SIGNAL 1 39
 FT CHAIN 40 747
 FT DOMAIN 448 542 FIBRONECTIN TYPE-III 1.
 FT ACT_SITE 546 639 FIBRONECTIN TYPE-III 2.
 FT ACT_SITE 208 208 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 349 349 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 747 AA: 78936 MW: 854730 DB54730 DB84280 CRC64;

Query Match 8.8%; Score 487; DB 1; Length 747;
 Best Local Similarity 27.1%; Pred. No. 7,9e-20;
 Matches 168; Conservative 67; Mismatches 175; Indels 210; Gaps 23;

OY 498 KFGDLQVNTGSSS-----VDLST-----VTYRWFTRD 525
 DB 260 KRGVFWTSTGLDYDGTWGMGNLRCVRHPIDLAGHODLVYSPHDYGPLVEFDQKWFQKD 319
 OY 526 GGSSTL--VYNCDMAAGCCNIR---ASFQSVNATPTATYTL-----QLSTF 568
 DB 320 FPKASLTLDVWGPNNMLFHDEIDIALPLIGEWGRUGODRQKMAALRLDLVAERLSOT 379
 OY 569 GCTL--AAGSGTEIONRVKSDMSNFTETNDYSYGTNTFFODWTKVTVV-----NG 619
 DB 380 FVWLPNNSDGTGGL--LDDMKTMDEV---KSTMLEPTLMKHGKGVGDHQPVLGG 432
 OY 620 V-----LVWGEDVTPPSVPTGLVYTVGVSSSVLANAST--NVGAHYNVR 667
 DB 433 VSTTGTSISQGGGTPPTAPTAPTAGTCTPTASTVPLTWSASPTDGGSGVAGYEYR 492
 OY 668 NGVILGQPTVSTFPTDGLAGTAYTYTAAVDAGNTSPRPVDCPTPPNONGVTSVD 727
 DB 493 GTTLTGTTATISYTVGLAADSAYTFVRAKAGNTS-----531
 OY 728 GEYVQTNEMNSAOCCLINTATGAMTVSTANFSGGTGACATPYPIKGCWGNCTTK 787
 DB 532 -----ASAATVARTA-----AGCGVATASVPTGLAGTPTAT 564
 OY 788 NKGMPLOISQISAVTSMSTVYSSGCAVDYADIMTNSPTTGGNGTEIMIMNSRG 847
 DB 565 SVPL-----TWASTDTGGSGVTGEYVKGSS--TLVAPRTGT-----599
 OY 848 VQFQSQTATGTVAGHTMNVWNGGOTSMTKISYVLTPEATSSINDLKAIFADAARGS 907
 DB 600 -----SHVYTGLSAA-----TAYTETVRA-----VDA--GN 624
 OY 908 INTSDYLLDVEAGFETWOGGGLGNSFSVSYSVSGVACRATVYVNSDMGSGFTATVYT 967
 DB 625 VSAASAPVCV-----TTAPDPTTG--SCAVITYTANG--WSGFTAVITLT 665
 OY 968 NKGSAATNMTYVAMSGNGQYVNTMNTALTOSGASVATINISNNVIOPGOSTTGFENG 1027
 DB 666 NTGTALTSMTLGFAPFSCQTLTGQMSARMAQSGSVTATNEMNANVLAPGASVEIGFSG 725
 OY 1028 SYSGTNAAPT-----LSCST 1041
 DB 726 THTGTNTATPATFTVGGATCT 745

RESULT 8
 GUNA_CALSA STANDARD; PRT: 1742 AA.
 AC P22534;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
 DE (Cellulase A).
 GN CELA.
 OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
 OC Bacteria: Firmicutes: Bacillus/Clostridium group;
 OC Thermoaerobacter group; Caldicellulosiruptor.
 OX NCBI_Taxid=44001;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95336703; PubMed=7612247;
 RA Te'o V.S., Saul D.J., Bergquist P.L.;
 RT "CellA, another gene coding for a multidomain cellulase from the
 RT extreme thermophile Caldocellum saccharolyticum.";
 RL Appl. Microbiol. Biotechnol. 43:291-296(1995).
 RN [2]
 RP SEQUENCE OF 1516-1742 FROM N.A.
 RX MEDLINE=91247819; PubMed=2039230;
 RA Leuht E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
 RT "Cloning, sequence analysis, and expression in Escherichia coli of a
 RT gene coding for a beta-mannanase from the extremely thermophilic
 RT bacterium 'Caldocellum saccharolyticum'.";
 RL Appl. Environ. Microbiol. 57:694-700(1991).
 CC -1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
 CC ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
 CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
 CC CELLULOSE.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- PTM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
 CC SITE FOR PROTEOLYSIS.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
 CC E (FAMILY 9 OF GLYCOSYL HYDROLASES).
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
 CC L (FAMILY 48 OF GLYCOSYL HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 DR EMBL: I32742; AAA91086.1; -;
 DR EMBL: M36063; AAA2860.1; -;
 DR EMBL: L01257; -; NOT_ANNOTATED_CDS.
 DR PIR: A43745; A43745.
 DR HSSP: P26221; TTF4.
 DR InterPro: IPR001956; CHD_3.
 DR InterPro: IPR000556; Glyco_hydro_48.
 DR InterPro: IPR001701; Glyco_hydro_9.
 DR Pfam: PF00942; CBD_3; 3.
 DR Pfam: PF02011; Glyco_hydro_48; 1.
 DR Pfam: PF00759; Glyco_hydro_9; 1.
 DR PRINTS: PR00844; GHYDRASE48.
 DR ProDom: PD001947; CBD_3; 2.
 DR ProDom: PD01903; Glyco_hydro_48; 1.
 DR ProDom: PD00592; Glycosyl_HYDROL_F9; 1.
 DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9; 1.
 DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9; 2; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
 FT SIGNAL 1 23
 FT CHAIN 24 1742
 FT DOMAIN 24 642
 FT DOMAIN 643 700
 FT DOMAIN 701 857
 FT DOMAIN 858 903
 FT DOMAIN 904 1060
 FT DOMAIN 1061 1112
 FT DOMAIN 1113 1742
 FT ACT_SITE 396 396
 FT ACT_SITE 434 434
 FT ACT_SITE 443 443
 FT CONFLICT 1545 1545
 T -> A (IN REF. 2).

SEQUENCE 1742 AA: 193696 MW: 3f0699a2123BED07 CRC64:

Query Match 8.1%; Score 449.5; DB 1; Length 1742;
Best Local Similarity 21.5%; Pred. No. 2.8e-17;
Matches 249; Conservative 114; Mismatches 389; Indels 407; Gaps 48;

```

OY 29 AGVLAQVSIASIVPLAMQHPAIAATV-----DNPAGATFFVNY-- 71
DB 173 AYVAETAAASIAIYLKRNPTKATYLOHAKDLYEAEVTKSDSGTAAANGYNSMSG 232
OY 72 -----WAOEVOSEANQINATLAKMRVSTYSTAVMMDRIAANGVGGPGLTYYIDA 125
DB 233 FYDELSMA-----AVHL--YLATND-----STYLYLK 256
OY 126 ALSQOQTEPEVIEIYIDLPGHCAALASGE--LPATLAOLTYEQYIDPISILSN 183
DB 257 AESVYQ-NMKRIGSNIDYKMAHCWDDVHNGMALILAITDKDYK-OLISHLDYWT 314
OY 184 P-----KYSRLAVTIEPDSLPNAVTVMSI-----QACATAV-----PYEGGLEY 225
DB 315 GYNGERIKYTPKGLAWLDQMSLRATTTFLAFVYSDMSGCTGKETRRKFGESQIDY 374
OY 226 AL-----TKLAIIPNYIYMDAHSQV-----LGMWNNAGCYVO 259
DB 375 ALGSTRGFVYGFCTNPPKRRHRTAHSWADSQSIPTSYHRHTLYGALVGGPSDDSYTD 434
OY 260 EVQKVLNASTVNGIDGFYNTANY-----TPLEKPFM----- 292
DB 435 DISNYVNEVACDYNAGVGLAMKYLIGGNPIPDFAKTEPTNDEFVEACINAGTGN 494
OY 293 -----TATQOVQGVESANFYQNNPDIDEADVADLYSRLVANGF-PSSIGMLIDTRLN 346
DB 495 FLEIKAIYVNOGGMARAPNKLKFR-----YFVDL-SELIXKAGYSPNDL-----TLST 541
OY 347 GNGGPNETGP-----STATDY-----NTPVNOCKIDLROHRLG-WCNQON 385
DB 542 NYNOGAKVSGPYVMDSSRNIIYILVDFTGTLIYPCGQOKKKEVOFRLAQAQVOMNSN 601
OY 386 G-----AGLGPQPAQSPDFPNALHDAAVYKIPGSESDGISAAS-DPTTGKKDPMQDP 438
DB 602 DYFQDIKGVSSGSVVKTYIPLDYEDIDIVW-----GEEPGISGVSPTPASVTPPTPTPT 657
OY 439 TYTTSYGLTALPNSPIAGQMPFAQFDLVANARPA-----VVSGLKQVYKNNDSA 491
DB 658 TATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 717
OY 492 PCGNQIKRGLQVNTGSSVDLSTVTVRYVFTFRDGGSTLVLYNCMDAAMCGCNIAASFGS 551
DB 718 STTWTIRPMLKLVNNGSSSIDLSRVTIRVWYTVDGERRAQSALS-DWAQIGASNVTFKFEVK 776
OY 552 VNPAPTADTYLQLSFTGGT--LAAGSGCEIONRVNKSMDSNFTENDSY-GINTPEQ 608
DB 777 LSSSVSGADYYLEIGFKSGAGQLOPGKDGEIOIRFKKDMSNVNOGNDMSWQMSWTSYSG 836
OY 609 DMTVTVVGVNGLVWGTEDVTPPSVPTGLVTVGVSSSVSLANNASTDNVGAHVYVYRN 668
DB 837 ENKEKTAIYDVLWVGQE-----PSGTTPA----- 861
OY 669 GVLVGOPTVTSFTDTGLAGATVYTYAADAAGNTSAPSPVDCRPGNONGVTSYODG 728
DB 862 -----PSTP-----TVYVTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 899
OY 729 EXRVQTNEMNSAQCCLINTATGAMTVSTANFSGGTGAGATYPSIYKGCCHNCCTKN 788
DB 900 -----PVST-----PAT----- 906
OY 789 VCMPIQISOIGSAVTSWSTTOVSSGAYVAY-DIWTNSTPTTTCOPNGCTEIMWLNSRG 847
DB 907 -----GGQIKVLANKETNSTNT-----IRPMLK----- 931
OY 848 VQFPGSQTATGVTAGHTVNNVAGQOOTSWKIISYVLPFGANSISMDLKAIFADAANGS 907

```

```

DB 932 VYNSGSSSIDLSRVTIRVWYTVDGERRAQSALSDWAQI-GASNVT-----EFYVLISS 983
OY 908 INTSDYLLDVEAGEETIOWGGGGLGNSFSVTSQVACRATRYVNSDMGSGFTATVTVT 967
DB 984 VSGADYYLEI--GFRKSGAGLOPGKDTGEIOIRN-----KQDMNSN-----Y 1023
OY 968 NNGSRATNCTVYAM-----SFGGNQTVTNWMTALTOSGASVYTNLXSNVIOPGOST 1022
DB 1024 NQG-----NDM--SWIOSMSTYGENEKVT-----AYIDGVLWVGQEP- 1058
OY 1023 FGFNGSYSGTNAAPTLCST 1041
DB 1059 -----SGATPAPVTVPT 1070

```

```

RESULT 9
MANB_CALSA STANDARD; PRT: 1331 AA.
ID MANB_CALSA
AC P22533;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta-mannanase/endoglucanase A precursor [includes: Mannan endo-1,4-
DE beta-mannosidase A (EC 3.2.1.78) (beta-mannanase) (Endo-1,4-
DE mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)].
GN MANA.
OS Caldocellum saccharolyticum (Caldicellulosigrupor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermotomacrobacter group; Caldicellulosigrupor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93119139; PubMed=1476429;
RA Gibbs M.D., Saul D.O., Luthi E., Bergquist P.L.;
RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a
RT multidomain enzyme."
RL Appl. Environ. Microbiol. 58:3864-3867(1992).
RN [2]
RP SEQUENCE OF 1-346 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Luehl E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
RT bacterium 'Caldocellum saccharolyticum'."
RL Appl. Environ. Microbiol. 57:694-700(1991).
CC -1- FUNCTION: DEGRADATION OF HEMICELLULOSES. THE SECOND MOST ABUNDANT
CC POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH
CC MANNANASE AND ENDOGLUCANASE ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
CC linkages in mannans, galactomannans, glucomannans, and
CC galactoglucomannans.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES
CC CELSIUS.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC J (FAMILY 44 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
CC EMBL: L01257; AAA71887.1;
CC EMBL: M36063; AAA72861.1;
CC PIR: B43745; B43745.
CC PIR: A48954; A48954.
CC HSSP: Q06851; INBC.

```


FT DOMAIN 462 562 CELLULOSE-BINDING (BY SIMILARITY).
 FT ACT_SITE 203 203 PROTON DONOR.
 FT ACT_SITE 323 323 NUCLEOPHILE.
 FT DISULFID 75 161
 FT DISULFID 209 212
 SO SEQUENCE 562 AA: 60747 MW: 848625640635041 CRC64:

Query Match 7.5%; Score 414; DB 1; Length 562;
 Best Local Similarity 22.8%; Pred. No. 6e-16;
 Matches 196; Conservative 80; Mismatches 220; Indels 364; Gaps 36;

227 LKTHAIIPNYIYDAHSGMLGPNNASGYOEVOKVLNASIGVNGIDGFTNTAN--- 283
 24 LAIYALANLAVPRAPRAAGGYW--HTSG--REILDANNVPRAGIMWFGEFCNVV 79
 284 -----YPLKEPPHMTAQVGVGPESANFVOWNPDIADAVDL 324
 80 HGLMSRDYRSMDOIKSLGYNTIRLPYSDILKPGTMP-NSINFYOMNDL-QGLTSLQV 137
 325 YSRIVAGFPSSIGMLDILRNGMGHPNPTGPSTADVTNFEVNSKIDLRHRLGIMCNO 384
 138 MDKIVA--IAGQIGL-----RIILDRHR----- 158
 385 NGAGLGPPOASPTDFPNALIDAYWIKPPGESDGTSAADPTTKKSDPCPTTTSY 444
 159 -----PDCSCGSAIMW-----TSVSEAT----- 177
 445 GVLINALPNSPIAGOWPFAOFDOLVANAAPVYSGGLKVOYKNNDSAPGDNQIKPGIQV 504
 178 -----W-----ISDQALAQ-----RKGNPTV-----VGFDLH 201
 505 NTGSSSVDLSTVTVRYWTFRDGSSITLVYNCNW--AAMCGNIFASGVSVP----- 554
 202 NEPHDPA-----CMCGGDP-----SIDWRLAARAGN--AYLSVAPNLLIEVEG 243
 555 -ATPPADRYLQSLFTGTLAAGSGTGEIQNRKNSDNSEFETDYSYGTTFQDMTKV 613
 244 VQSYNGDST---WMGNIQAGQYPPVLANVFNRLVYS---ANDYA-----T 283
 614 TVYVNGVLWCTEDVTPSPVPTGLVYVGVSGSSVSLANNASTDNCVANHVNRYNGVLV 673
 284 SYVPO---TWFSDPTEFPNMP-----GIMNKNMGYLENQNAPVWLEFET----- 326
 674 QPVTYSFTD-TGLAGTATYTVAAVDAAGNTSAPSTPVDCTPGPNQGVTSVODGEYRV 732
 327 --FLQSTTDQWTKLVOLRPPA-----QYGADSFQ----- 356
 733 QTEMNNSAQQCLTINTATGAMTVSTAN-FSGGTGAPATYTSYKGGHMGCTKKNYG- 790
 357 -----WTFMSMNPDSGDTG-----ILKD-DMQTVTVKDOY 387
 791 -WPIQIS---QIGSAVTSMTQVS---SGAYDAVDIWNTPPTTGCPNGCTEIMILN 843
 388 LAPKKSIFDPVAGASAPSSQSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 438
 844 SRGVOFGSGQFATGVTVAGHTNVMVQGOQTSKIIISYVLTGCAISISNLDLKAIFADA 903
 439 -----PTLIPPAIPPTTAPSP-----TPSPPTAAS----- 461
 904 ARGSLMTSDYLLDVEAFELIMOGGQGLGSSFSVYSGVACRATYVNSDMGSGFTAT 963
 462 -----GARTASIQVNSDMGNGFTYV 482
 964 VTVNTGSRATNGTVAMSGFQNOTVTNYNTALTQSGASVATNLKSNVIOPGOSTTP 1023
 483 VAVYNSGSAKTKTWTSMTFEGNOTITNSMNAVATQNGOSVATAMSYNNVIOPGONTTF 542
 1024 GFNGSYGCTNAAPTLISCTAS 1043
 543 GFQASYSCTGSMNAAPVACAAS 562

RESULT 11
 CIPALCIOTM STANDARD: PRT: 1853 AA.
 ID CIPALCIOTM 006851;
 AC 01-JUN-1994 (Rel. 29, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Cellulosomal scaffolding protein A precursor (Cellulosomal glycoprotein S1/Sr) (Cellulose integrating protein A) (Cohesin).
 GN CIPa.
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 ON NCBI_TaxID=1515;
 RX SEQUENCE FROM N.A., AND SEQUENCE OF 29-40 AND 42-43.
 RC STRAIN=ATCC 27405;
 RX MEDLINE=9320508; PubMed=8316083;
 RA Gerngross U.T., Romaniec M.P.M., Kobayashi T., Huskisson N.S.,
 RA Demain A.L.;
 RT "Sequencing of a Clostridium thermocellum gene (cipa) encoding the
 RT cellulosomal Sl-protein reveals an unusual degree of internal
 RT homology";
 RL Mol. Microbiol. 8:325-334(1993).
 RN [1]
 RP SEQUENCE OF 1820-1853 FROM N.A.
 RX MEDLINE=93209931; PubMed=8458832;
 RA Fujino T., Beguin P., Aubert J.-P.;
 RT "Organization of a Clostridium thermocellum gene cluster encoding the
 RT cellulosomal scaffolding protein CIPa and a protein possibly involved
 RT in attachment of the cellulosome to the cell surface.";
 RL J. Bacteriol. 175:1891-1899(1993).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.
 RX MEDLINE=97238934; PubMed=9083107;
 RA Shimon L.J.W., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y.,
 RA Frolow F.;
 RT "A cohesin domain from Clostridium thermocellum: the crystal
 RT structure provides new insights into cellulosome assembly.";
 RL Structure 5:381-390(1997).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.
 RX MEDLINE=97076134; PubMed=8918451;
 RA Torno J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,
 RA Steltz T.A.;
 RT "Crystal structure of a bacterial family-III cellulose-binding
 RT domain: a general mechanism for attachment to cellulose.";
 RL EMBO J. 15:5739-5751(1996).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.
 RX MEDLINE=98022914; PubMed=9402065;
 RA Tavares G.A., Beguin P., Alzari P.M.;
 RT "The crystal structure of a type I cohesin domain at 1.7-A
 RT resolution";
 RL J. Mol. Biol. 273:701-713(1997).
 RL J. FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
 CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
 CC CELLULOYTIC ENZYMES.
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE.
 CC -1- DOMAIN: CONTAINS 9 COPIES OF A DOMAIN (COHESIN) OF ABOUT 150
 CC RESIDUES THAT BINDS TO THE DOCKERIN DOMAIN BORN BY THE CATALYTIC
 CC COMPONENTS OF THE CELLULOSE.
 CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
 CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
 CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```
CC -----NOT_ANNOTATED_CDS.-----
DR EMBL: X67506; CAA47840.1; -.
DR PIR: S36859; S36859.
DR PDB: 1ANU; 23-JUL-97.
DR PDB: 1AOH; 08-JUL-98.
DR PDB: 1NBC; 26-SEP-97.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR002102; Cnhesin.
DR InterPro: IPR002105; Cnhesin_1.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00963; Cnhesin; 9.
DR Pfam: PF00404; Dockerin; 1; 2.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE: PS00448; CLOS_CELULOSOME_RPT; 2.
DR Cellulose degradation; Cell wall; Glycoprotein; Repeat; Signal;
KW 3D-structure.
FT SIGNAL 1 28
FT CHAIN 29 1853 CELLOSOMAL SCAFFOLDING PROTEIN A.
FT DOMAIN 364 522 CELULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 54 1694 9 X 150 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 29 182 1-1.
FT REPEAT 183 322 1-2.
FT REPEAT 560 704 1-3.
FT REPEAT 724 866 1-4.
FT REPEAT 889 1031 1-5.
FT REPEAT 1054 1196 1-6.
FT REPEAT 1219 1361 1-7.
FT REPEAT 1384 1526 1-8.
FT REPEAT 1548 1690 1-9.
FT DOMAIN 1791 1846 2 X 24 AA APPROXIMATE REPEATS.
FT REPEAT 1791 1814 2-1.
FT REPEAT 1824 1847 2-2.
FT COMPLET 1615 1615 A -> AA (IN REF. 1).
FT SEQUENCE 1853 AA; 196831 MW; 3ABDDC03ABRC572 CRC64;

Query Match 7.5%; Score 413; DB 1; Length 1853;
Best Local Similarity 22.3%; Pred. No. 3.2e-15;
Matches 263; Conservative 128; Mismatches 363; Indels 424; Gaps 59;

OY 119 LTTVDAALISQGGTPPEIEIYVDLPGRCAALASNEIPATAGL----- 166
DB 14 LTTTAAAMPVVSAAATVET-----GKTAAGSKVEIPITLKGVPSKGMANCEVYL 67
OY 167 ---QVETQYIDPIASILSNPKYS-----LRTVITIEPDSLPN--AVTNMSI- 209
DB 68 GYDPVAVLETEVKP--GSIHKDPPSKSPDSALYPRKMKVLFPAEDSGGTAITQDGVF 126
OY 210 -----QACATAVPYEAGEIEVALTKLHAIPNVIY--MDAAHSGMLGMPNNAS-GYV 258
DB 127 ATIVATVKSAAAPITLLEVG-AFADNDLVEISTFVAGVNLGSSVPTQNPVSDGVV 185
OY 259 QEVQVVLASAGVNIIDFVTNTANYTPLKKEPMTATQVQGPYESANF-YQAMPD--- 314
DB 186 VEIGKV-TGSVG-----TVEIP--VFYKGVPSKGIANDVEFRYDPNVLE 228
OY 315 ---IDEADVAVDLYSRVLAAGFPSSIGMLIDTLRMGWSGPNPTGPTSVATDVNTFVNOGK 371
DB 229 IIGIDPGLIYD-----PN-----PTYSFDTAIPDKRI 257
OY 372 IDLRQHRIGIMCNGAGLGQPPQASPTDFPNAHLDAYWIKPPG-----ESDGTSAAD-- 425
DB 258 IVF-----LFAEDSGTG---AYAITKDGVPFAKIRATVKSAPGYITTEDEVGAFDNDLV 308
OY 426 -----PTTGKSPKCDPYT-----TSYVLNALPNSIACQMPFA 463
DB 309 EOKVSFIDGGVNVGNATPKGATPNTATPTKSATATPRPSVPTNTPTNTP----- 360
OY 464 QFDQLVANAARPAVVSGLGVQYKKNDSAPGDNIKPGLOLVNTGSSVDLSVTVRYWPT 523
DB 464 QFDQLVANAARPAVVSGLGVQYKKNDSAPGDNIKPGLOLVNTGSSVDLSVTVRYWPT 523
```

```
DB 361 -----AN---TPVSGNLKVEEYNSNPSTINSINPQKRVNTWGSSAIDLAKTLIRYTT 411
OY 524 RDGGSSTIYNCDMAAM-----GCCNIRASGVSVPATPATDYLOLSEFTGTLAA 574
DB 412 VDQKDDQTFW-CDHAALISNGSYNGITSNVKGTGVKKSSSTNNADTYLEISFTGTLLEP 470
OY 575 GGSSTGEIQNRVAKSMDNSNTEFNDSYGNITTFQWTKYTVVNVGLWGTED----- 627
DB 471 GAHV-QIOGRFAKNDSMNSYTSQSDYSFKSASQFVEMDQVATLNVGLWKGKGGSVVPS 529
OY 628 ---VTPSV-----PTGL-----VTVSSSSVSIAMNAS-TDNY 658
DB 530 TQPVTPPATTPPATTPPATTPPATTPPATTPPATTPPATTPPATTPPATTPPATTPPATTP 589
OY 659 GVAHYN-VY-----RNGVLYGQPTVTSFTDGL-----AAGT-AY 691
DB 590 GIANDCFVSYDPNVLEIIEIKPGLIYDPNDKSFDFAVYDPDKRIYVLFPAEDSGTAGY 649
OY 692 TYT-----VAAYDA-----AGNTSAPS 708
DB 650 AITKDGVEATIVAKKSGAPNGLSVIKFEVVGAFANNDLVEORTQFPDGVNVGDTTVPPT 709
OY 709 TPVD--CFPGPNQNGVTSVQDEYRQVOTENMNSAQQCLTINTATGAMTVSTANPSGGTG 766
DB 710 TTTTVPVTPPTDSNAV-----RIKVD-----TVAKFGDVIYRIYVRESG--- 748
OY 767 GAPATPYSIYKG---CHWGCNCTKNVGMFIQISQISAVT-----SMST----- 808
DB 749 -----IPS--KGIANDCFVSYDPNVLEIIEI-EPGDIVDPNDPKSFDLAVYDPDKRIY 800
OY 809 -----QVSSGADVAYD--IMTNSPTTGGPNGEIMIMLSRG-----GVOPPG 852
DB 801 FLEAFDSSGTGAATIKKDGVEATIVAKKSGAPNGLSVIKFEVVGAFANNDLVEORTQFED 860
OY 853 SQTATGVYVAGHVMWVWQOQTSWKIISVLTTPGATPISNLDLKA1----- 898
DB 861 GGVNVGDTTTPAT-----PTTPVTPPTTDDLDVARIKVDIVNAKPGDT 904
OY 899 -----PADAAAGSLNTSDY-----LLDVEAGEIEIWGGOGGLGNSPSVSV--- 939
DB 905 VRIPVRFSGIPSKGIAN-CDFVSYDPNVLEIIEIEPDGIYDPPN--DKSFDIAYVDP 960
OY 940 -----TSGVACRATYVNSDMGSGFTATYTVNTGSRATNGWTV-----AMSG 984
DB 961 KRIVFLEAFEDSG---TGAVAITRD--GVFATL-VAKVKSAGAPNGLSVIKFEVVGGA 1012
OY 985 GNOQTVNTWNTALTOGASVATATNLSNVNIOPGSTT 1022
DB 1013 NNDLVEQ--KIQFPGGVNVGDTT-----EPATPTT 1041

RESULT 12
CIPB_CLOTM
ID CIPB_CLOTM STANDARD; PRT; 772 AA.
AC 001866;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cellulosomal scaffolding protein B (Cellulosomal glycoprotein SI/SL)
DE (Cellulose integrating protein B) (Fragment).
GN CIPB.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillae/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS;
RA MEDLINE=93146373; PubMed=1490597;
RA Poole D.M., Morag E., Lamed R., Bayer E.A., Hazlewood G.P.,
RA Gilbert H.J.;
RT "Identification of the cellulose-binding domain of the cellulosome
subunit SI from Clostridium thermocellum YS.";
```


QY 759 ANPSGCTGAPATYPSIYKCHMGCTYKN-----VGMPIQISIGSAVTSMS 806
 DB 651 NMFYGTG-----TWVFGV-W-SCNKLKDTLAEIANRGFNLRVPS-----AELILNM- 697
 QY 807 TTVQSSGAY---DVAIDITNSTPTTGTGPN-----GTEIMIMLNSRGVQ 849
 DB 698 ---SOCIYPRPNINYV---NPELEGKNSLEVPDIYVQCKEYGLKIMLDIHS----- 744
 QY 850 PFGSQTATGTVAGHTMWNWQGOOTS----- 875
 DB 745 -----IKTDAMGHIPYVYDEKFTPEDEYKACEWITRNRKNDITIAFDLKNBPGKP 797
 QY 876 WKIISYLVTEGATISIMULDKAIFADAANGSLNSTDYLLDVEAGFE----- 922
 DB 798 WQDPTFAKWNSTDINMKYA---AETCAKRIINPNLLIYEGIEAYPKDDVTWTSKS 854
 QY 923 -----TWGGGGLGNSFSYSVTSGCVACRATY-----VYNSDW-GSGFTATVTV 967
 DB 855 SSDYSTWVGSLNLRGVKRPYINL--GKYQKVVYSPHDYGPVYQOPWYEGFTKE----- 908
 QY 968 NTGSRATNGWTVANSF 983
 DB 909 ---SLQDCMRPNMAY 921

RESULT 15
 GUN2_BACSU STANDARD: PRT: 499 AA.
 ID GUN2_BACSU
 AC P10475;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 GN (Carboxymethyl-cellulase) (CMCase) (Cellulase).
 DN BGIC OR GLD OR EGLS.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-PAP115;
 RA MEDLINE=87066783; PubMed=3024130;
 RA Mackey R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
 RA Moraneill F., Seliy V.;
 RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene";
 RT Nucleic Acids Res. 14:9159-9170(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CK-2;
 RA MEDLINE=95225656; PubMed=7710280;
 RA Lindahl V., Aa K., Tronsmo A.;
 RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
 RT subtilis CK-2";
 RT Antonie Van Leeuwenhoek 66:327-332(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA Rose M., Entlan K.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 30-45.
 RC STRAIN-CK-2;
 RA MEDLINE=95225655; PubMed=7710279;
 RA Aa K., Flengstrand R., Lindahl V., Tronsmo A.;
 RT "Characterization of production and enzyme properties of an
 RT endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
 RT compost soil";
 RT Antonie Van Leeuwenhoek 66:319-326(1994)
 RL -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL

HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Z29076; CAA82317.1;
 DR EMBL: X04689; CAA48392.1;
 DR EMBL: X67044; CAA47429.1;
 DR EMBL: Z73234; CAA97610.1; ALT_INIT.
 DR EMBL: Z99113; CAB13696.1; ALT_INIT.
 DR PIR: A26114; A26114.
 DR HSSP: O85465; 1A3H.
 DR Subtilist; B610437; bgIC.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR001547; Glyco_hydro_F5.
 DR Pfam: PF00942; CBD_3; 1.
 DR Pfam: PF00150; cellulase; 1.
 DR ProDom: PD001947; CBD_3; 1.
 DR ProSite: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR Cellulose degradation; Hydrolyase; Glycosidase; Signal;
 KW Complete proteome.
 FT SIGNAL 1
 FT CHAIN 30 499
 FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
 FT CONFLICT 283 283 S -> N (IN REF. 2).
 SQ SEQUENCE 499 AA: 55287 MM; 8F735FEF71B3EAE2 CRC64;

Query Match 6.9%; Score 382; DB 1; Length 499;
 Best Local Similarity 26.0%; Pred. No. 3e-14;
 Matches 133; Conservative 69; Mismatches 154; Indels 156; Gaps 24;

QY 164 AGIQTEYOTYIDPIASILSNPKYSS-----LRVITIE-----PDSLNNVNTMS 208
 DB 93 AAMYTADGCTID-----NPSYKNKYKAEVAKELGIYIIDHMLDGNPN----- 139
 QY 209 IQACATVAPYEEQIEIETLTKLHA-IPNVIYMDAASHGLGNPNASGYOQKVLNA 267
 DB 140 -QNEKKAKEFFE-----MSLYGNPNVIYEIANPNNGDVNKKRIKPYAEVLSVILK 193
 QY 268 SIGVNGIDGIV--TNTANTPLKEPNTATQGVGQPVESANFYQNPDI-DEAD--- 319
 DB 194 ---NDPNNIIIVGTGT-----WSDVDNDAADQLK 220
 QY 320 ---YAVDLSRLVAAFPSSIGMLIDTLRN--GMGFPNEPT--GPSTAT-DVNTFVN 368
 DB 221 DANWYALHFFA-----GTHGQPLRDKANYALSKGADIFVTEWGTSDASNGGVFLD 272
 QY 369 QSKIDLR--QHRGLMCNONGAGLGQPPQASPTFPNAHLDAYVIKPPESDGTSAASD 425
 DB 273 OSREWLKYLDSKTSISWVNN--LSDKOESSA-----LKPAGSKTGMRLSD 317
 QY 426 PTT-----GKSDPMCDPTYTTSYGVTLNALPNSPIAGQWPAQDQLVANRPA 475
 DB 318 LSAAGTFVRENTLGTKDS-----TRDIPETPSKDR--PTQ----- 350
 QY 476 VVSGGLKVOYKNNDSAPGDNOIKPGLQLVNTGSSVDSVSTVTVRYWF--TRDGSSTLVYN 534
 DB 351 --ENGISVOYRAGDGSMSNQLRPOLQIKNNCNTVYDLKDVYARWYKAKKNGN---FD 405
 QY 535 CDMAAMCGCNIRASFESVNPATPTADYVQLQSLFTGCTLAAGSGEIONRNKNDWMSFT 594
 DB 406 CDYAQIGCGGVTHKFEYTLKPKQAGDYDELQFKKGTLPAGSTGNQLRLHNDWMSNYA 465
 QY 595 ETDYSGTNTFTQDPTKTYVYVNGVLVWGTE 626

DB 466 QSGDYSEKSNTEKTTKTTLYDQSKLIWTE 497

Search completed: August 29, 2002, 16:38:45
Job time: 965 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:37:47 ; Search time 125.36 Seconds
(without alignments)
1439.324 Million cell updates/sec

Title: US-09-917-384-6
Perfect score: 1 NERTQSGRNCRCYRGRTTM.....GFGSYSGTNAAPTLSCTAS 1043
Sequence:

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:
2: SP bacteria:
3: SP fungi:
4: SP human:
5: SP invertebrate:
6: SP mammal:
7: SP mhc:
8: SP organelle:
9: SP phage:
10: SP plant:
11: SP rodent:
12: SP virus:
13: SP vertebrate:
14: SP unclassified:
15: SP viirus:
16: SP bacteriap:
17: SP archaeap:

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1239.5	22.4	683	16	Q9PDW2
2	1051	19.0	579	2	086730
3	1015.5	18.3	454	2	09KH72
4	1000	18.1	596	2	060029
5	722.5	13.1	351	2	09S3V4
6	705	12.7	490	3	096V97
7	694	12.5	491	3	096V98
8	649	11.7	377	2	008468
9	647.5	11.7	381	2	09RJY3
10	633.5	11.4	371	2	054331
11	629.5	11.2	382	2	09KH11
12	618.5	11.2	460	3	059963
13	597.5	10.8	444	3	002321
14	589.5	10.6	443	3	096V02
15	581.5	10.5	436	3	09C1R4
16					Q9P8N1

17	569	10.3	465	3	Q96TP4	Q96TP4 pleurotus s
18	565.5	10.2	476	3	Q9C1S9	Q9C1S9 humicola in
19	554.5	10.0	384	2	Q9X602	Q9X602 streptomyc
20	554	10.0	471	3	Q9X602	Q9X602 streptomyc
21	544	9.8	457	3	Q9X8Y8	Q9X8Y8 trichoderma
22	499	9.0	472	3	Q93860	Q93860 acromonium
23	495.5	9.0	930	2	Q9REX5	Q9REX5 piromyces r
24	482.5	8.7	921	2	Q9RLB8	Q9RLB8 caldicellul
25	457	8.4	1751	2	Q9AOC4	Q9AOC4 caldicellul
26	452	8.2	458	3	Q9UW11	Q9UW11 piromyces r
27	450.5	8.1	1711	2	Q9X3P6	Q9X3P6 caldicellul
28	449	8.1	1426	3	P78721	P78721 orpionmyces
29	444.5	8.0	460	3	Q9UW10	Q9UW10 piromyces r
30	443	8.0	428	3	Q9X3P5	Q9X3P5 caldicellul
31	436	7.9	428	3	Q9X3P5	Q9X3P5 caldicellul
32	434	7.8	1770	2	Q9X3P5	Q9X3P5 caldicellul
33	431.5	7.8	1779	2	Q9X3P5	Q9X3P5 caldicellul
34	421	7.6	1915	2	Q9P808	Q9P808 piromyces r
35	417	7.5	376	3	Q9P808	Q9P808 piromyces r
36	412	7.4	260	3	P78720	P78720 orpionmyces
37	412	7.4	1000	2	Q24820	Q24820 thermophil
38	410	7.4	260	2	Q33897	Q33897 rhodothermu
39	407.5	7.4	2316	2	Q92411	Q92411 bacillus su
40	402	7.3	997	2	Q92411	Q92411 bacillus su
41	386.5	6.9	890	2	Q96D27	Q96D27 streptomyc
42	384.5	6.9	890	2	Q45532	Q45532 bacillus su
43	383	6.9	499	2	Q93LD0	Q93LD0 bacillus su
44	381	6.9	508	2	Q93TD6	Q93TD6 bacillus su
45	380	6.9	499	2	Q93TD6	Q93TD6 bacillus su

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	683 AA.
Q9PDW2	Q9PDW2			
AC	Q9PDW2			
DT	01-OCT-2000 (TREMBLREL. 15, Created)			
DT	01-OCT-2000 (TREMBLREL. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBLREL. 19, Last annotation update)			
DE	1,4-BETA-CELLULOBIOSIDASE.			
GN	XP1267.			
OS	Xyloella fastidiosa.			
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;			
OC	Xyloella.			
OX	NCBI_TaxID=2371:			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-9A5C:			
RC	MEDLINE-20365717; PubMed-10910347;			
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,			
RA	Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,			
RA	Bueno M.R.P., Bonaccorsi E.D., Bordin S., Boye J.M., Briones M.R.S.,			
RA	Colautino N.B., Colombo A.A., Camargo L.E.A., Carraro D.M., Carer H.,			
RA	Coulthart L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,			
RA	Faciociani A.P., Ferreira A.J.S., Ferreira V.C.A., Furian L.R.,			
RA	Fraga J.S., Franca S.C., Franco M.C., Fromme M., Furian L.R.,			
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,			
RA	Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,			
RA	Krieger J.E., Lemos M.V.F., Lopes S.A., Lopes C.R., Leite L.C.C.,			
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Leite L.C.C.,			
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Matsukuma A.Y.,			
RA	Marques C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,			
RA	Moore D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,			
RA	Nhai A.J., Nobrega F.G., Nunes L.R., Oliveira M.A., Paris A.,			
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Pesarin J.B.,			
RA	Pelajo R.B., Pereira P.G., Pereira H.A., Jr., Pesarin J.B.,			
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,			
RA	de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,			
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,			

RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.
RA da Souza A.P., Terezi M.F., Truffi D., Tassi S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Veljovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Melandis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"
RL Nature 406:151-159(2000).
RL EMBL; AEO03960; AAF84076.1; -.
DR HSSP; P07983; ICB2.
DR InterPro: IPR001524; glyco_hydro-6.
DR Pfam: PF01341; Glyco_hydro-6; 1.
DR PRINTS; PR00733; GHYDRLASE6.
DR ProDom: PD003733; glyco_hydro_6; 2.
KW Complete proteome.
SQ SEQUENCE 683 AA; 70852 MW; DCFB5F76B8E8D5F CRC64;

Query Match	22.4%	Score	1239.5	DB	16	Length	683
Best Local Similarity	31.9%	Pred	No. 9.8e-55				
Matches	314	Conservative	124	Mismatches	220	Indels	327
						Gaps	27

QY	53	AATHDNVYACATFEFVNYAEOSEJAEANOINTTLLAKKRVASTVETAWMARIAING	112
Dd	18	AEAHVNDNFVATISYLPNDISKDIDTSLAKVNDVSLAEKQOTIKNYTGTWGLDIDAVKG	77
QY	113	VNGGFG--LFTYIDAALISQOOGTTPPEVIEIYIDLPGRDCAALASNGELPATAAGLOTY	169
Dd	78	GPKNPGRLNTLGHDAALAKKGNTPITAFVYIDIGRCHALASNGELPLRPEGJORY	137
QY	170	EYOYIDPLASTLSNPKYSLSIAYVLIIEPDSLPAVNTWMSIOACATA--VYYEGDIEYAL	227
Dd	138	KREYIDITLASIFSNPKKDIRIVNLIBDLSPLNITWMSPPSCQAQANNTGYIEEGIKYAL	197
QY	228	TKLHAIPNVYIYMDAASHGMLGMPNNAAGYOEYOKVL--NASIVNGIDGFVNTANYTP	286
Dd	198	NKLSIPIPVNYMYMDIGHGWMGMOTNINLPAVSLTKYIOSTIAGFASVNFARINTANTTP	257
QY	287	LKEPPM-TATQOVGGOPEYASNFYQWMPDIDEADYAVDLSRLVAAGFPSSIGMLIDTLR	345
Dd	258	LIEPLMPNDLNINGOPFIRSSKFYEMNRYFEDMDYSETLVNDFVAAGWPSSIFIIDTGR	317
QY	346	NGMGPNRPPTSPATADIVNTFVNOISKIDLQRHGLKMGONAGIGAGORPOASPIDFPAHL	405
Dd	318	NGMGPEPRTS-APGNDVNSTVNSGRIDRRNHRKMNQADAGIGLPIAT---PGHV	372
QY	406	DAYVINKPGESEDGSAADPTTKKSDPMCDPYTTSYGVLTALNPSEIAGOMPEAFQ	465
Dd	373	DAFOIKRPPGVSDGSSSLIPMDQKGRPYCDPFETTPDGLTIALPDAISLGDWFIAGF	432
QY	466	DOLVANARPAYVSGLAKVOYRNNDSPADNOIKFGLDVLNTGSSVDLSIVYRYKWFTRD	535
Dd	433	VKLINNAPPDI-----SRTPPLP-----SGSSS-----S	456
QY	526	GSSSTLYVNCMDAAMGGINASGVSVPATPATDYTLQISFTGGLTLAGGSGTEIONRY	585
Dd	457	SGSS-----SSGGS-----SSSGSSSSSSGSSS-----	479
QY	586	NKSDMSNETENDYSYGTNTTFODMTKVTYVYVNGVLWGTEDVYPPSVPGLVYTVGSGS	645
Dd	480	-----SSGSSS-----SGSSSS	492
QY	646	SVSLAMNASTDNGVAHAINRYRNGVLVGOPIVTSFTDTGLAAGTAYTYTAAADACNTS	705
Dd	493	SGSSSSSSSSSSSG-----SSSSSSSSSS	515
QY	706	APSTVVDCTPGPNNGVTSVQDGEYRQTNEMNSSAQOCLINTANGAMVYRANPFGCT	765
Dd	516	SSGS-----SSSGSSSSSSSGSSSSSGS	537
QY	766	GGAPATYSIYKGGHMGNCCTKNVGMPIQISQISGAVTSMSTTSGGADVAVDIWTNS	825
Dd	538	S-----SSGSSSSSSSGSSS-----SSS	558
QY	826	TPPTTGGQPNTEIMIWLSRCGVOPFGSQATGVTAGHTWNVWGOQOTSMKIIISVYLP	885

[illegible]

RESULT	2	
086730		
ID	086730	PRELIMINARY;
		PRT;
		579 AA

```

DE      01-NOV-1998 (TREMBLrel.08, Created)
DT      01-NOV-1998 (TREMBLrel.08, Last sequence update)
DT      01-JUN-2001 (TREMBLrel.17, Last annotation update)
DE      PUTATIVE SECRETED CELLULASE.
GN      SC5C7.33.
OS      Streptomyces coelicolor.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX      NCBI_TaxID=1902;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-A3(2);
RA      Seeger K.J., Harris D.;
RL      Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-A3(2);
RA      parkhill J., Barrell B.G., Randalream M.A.;
RL      Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-A3(2);
RC      MEDLINE=97000351; PubMed=8843436;
RA      Redenbach M., Kleiser H.M., Denapalte D., Eichner A., Cullum J.,
RA      Kinash H., Hopwood D.A.;
RT      "A set of ordered cosmids and a detailed genetic and physical map for
RT      the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL      Mol. Microbiol. 21:77-96(1996).
DR      EMBL; AL031515; CAA20645.1; .
DR      HSSP; P07986; IEXG.
DR      InterPro; IPR001919; CBD.2.
DR      InterPro; IPR001524; glyco_hydro-6.
DR      Pfam; PF00553; CBD_2; 1.
DR      Pfam; PF01341; glyco_hydro-6; 1.
DR      PRINTS; PR00733; GLHYDRLASE6.
DR      PRODom; PD003733; glyco_hydro-6; 2.
DR      PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
SQ      SEQUENCE 579 AA; 60831 MW; 5A51DD28E7996A89 CRC64;

Query Match      19.0%; Score 1051; DB 2; Length 579;
Best Local Similarity 49.2%; Pred. No.2,4e-45;
Matches 215; Conservative % 61; Mismatches 131; Indels 30; Gaps

QY      50 PALATHTDNRPAAGTFFVFNRYNAQEOVSEANNTNATLAKKRVSTYSTAVWMDRIA 109
DB      158 PPGPGDRNDNPDPGAGQVYVNEPMANAAEFGGDR-----INDEPTGWLDRIAA 207
QY      110 INGVANGSGGLTFTYDAALSOOGGTPTEYIEIVYDLPGDCAALASNGELPATPAAGLQTY 169
DB      208 IEGANGSGGLRDHDHAL-EDKSGGENGVOLVITNLPGDSCSALASNGELGPE--IDRY 264
QY      170 ETQYIDTPASILSNPKVSSSLIVIIIEPDSLPNAVTVNMSIOACAT-----AVPYEQG 222

```

[illegible]

ID	09KH72	PRELIMINARY;	PRT;	454 AA.
AC	09KH72			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	CELLULASE PRECURSOR.			
GN	EX.			
OC	Thermomonospora fusca.			
OC	Bacteria: Firmicutes; Actinobacteria: Actinobacteridae;			
OC	Actinomycetales; Streptosporangineae; Thermomonosporaceae;			
OC	Thermobifida.			
OX	NCBI_Taxid=2021;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Al Y.-C., Wilson D.B.;			
RT	"Genomic DNA sequence encoding Thermomonospora fusca cellulase with overexpression."			
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AE268074; AAF75786.1; "			
DR	HSP: P07987; ICB2.			
DR	InterPro: IPR001524; Glyco_hydro-6.			
DR	Pfam: PF01341; Glyco_hydro-6; 1.			
DR	PRINTS: PR00733; GLYDRLASE6.			
DR	ProDom: PD003733; Glyco_hydro-6; 2.			
DR	PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.			
KW	Signal.			
FT	SIGNAL	1	31	POTENTIAL.
FT	CHAIN	32	454	CELLULOSE.
FT	SEQUENCE	454 AA;	48624 MW;	F8CD1C4CD4807134 CRC64;
Query Match		18.38;	Score 1015.5;	DB 2; Length 454;
Best Local Similarity		47.33;	Pred. No. 1.1e-43;	
Matches 221; Conservative		60; Mismatches 147;	Indels 39;	Gaps 12;
QY	27 LAGVAGVSTVAGVSTVAVMDRIALING-----VDNPFAGAFVFNVPYAGVQSFAN 82			
DB	6 LRR--LLGAAA--RALVSAALALFPGQAANNGEKVDNFEFGAKLYVNPVSAKNAAPGG 62			
QY	83 QTNATLAAFRVSVSTVAVMDRIALING-----VNGSGGLTYTIDALASQOQGTPEVI 138			
DB	63 S-----AVANSTAVWLDRIGALLEGNDSPPTGSMGLRDHLEAV--RQSGDPLTI 111			
QY	139 EIIVYIDPGKDCALASNGELPATAGLQTYETQYIDIASLISN--PYSSLRYVTIIP 197			
DB	112 QVATYINLPGKDCALASNGELPGE--LDRIYSEYVIDIADIMWFDAYENLRIVAIET 169			

Qy	198	DSLPLNATNMSTIOACATVPY-----YEOGIFVALFUKLHAIPNYIYMDAAHGWLAMP	251
		:	
Db	170	DSLPLNATYVGVNGSTETELCAATMKONGGVNGVGLRKLGELIPNYINIDAHHGWTGMD	229
		:	
Qy	252	NNASGYVOEYOKVLNAS--IGVNGIDGFEVNTANTYTLKEPEMTATIQOYGQGVESANFYQ	310
		: : : : : : :	
Db	230	SNRGSPVDFEYEAANASGTVDYVHGFEISNTANISATVEYEDVNGTNGQLIRGSKMYD	289
		: : : : : : :	
Qy	311	WNPDIIDAXVADYLSRLVAAAGFPSSIMLIDTLRNGMGGPNBPTGPSTATDVNTFVNOS	379
		:	
Db	290	WNOYVDLSTFYVDLRLALAKFRFSDIMLIDTSLRNGMGGPNBPTGPSSSTLDLNTYVDES	344
		:	
Qy	371	KIDLRQHRGLMGNOCAGICGCPPOKASPTPDPFNALHLDYVWIKPCCGSDGSAASPPTGK	431
		:	
Db	350	RIDRIRHFGWMCQAQAGAGERTVNP-----PGVDATVWYKPPGSDGASSEIPDECK	40
		:	
Qy	431	KSDPDCPTTYSIGYLIN---ALPNSPIAGWMPPAQFDQLVANARP	474
		:	
Db	406	GFDRKRCPTYGNNARNGNPNNSGALPAPAPISGHWFSAQERELLANAYP	452

Query Match	Best local Similarity	Score	DB 2:	Length	596:
Matches 210; Conservative	47.7%; Pred. No. 9.3e-43;	18.1%;	DB 2:	Length 596;	
	55; Mismatches 143; Indels 32; Gaps				
QY 50	PAIAAHVNPAPVAGATFEFNPYMAOEVSSEANQTNATLAAKMRVSTYSTAVMDRIIA 109				
DB 172	PTNPERGVNDPEFGAKLIYNPYMSAKAAAEPEGSS-----AVANESTAVWLDRIGA 221				
QY 110	ING-----VNGSGPLTITDALSQOQGTTPEVIEIYITDIDGRCDAALASNGELPATRAAG 165				
DB 222	IEGNDSPETGSMGLRDHLEAV-RQSGGDPILTIQVYITNLPERDCDAALASNGELGPDE-- 278				
QY 166	LQYIEYQIDPIASISLN-PKYSLSRIYTIIEPDSLPAVNTMSIQACATAPV----- 218				
DB 279	LDRKSEYIDIPADIMDFADYNLITVALIIEIDSLPNLVTYVNGNGTELCAYAKONG 338				
QY 219	YEQGIEFALTRILAIHPNVYIYMDAASGWLGMFNNAAGCYVQEVQVLAAS-IGVNGIDGF 277				

```
Db 339 YNNGVGYALRKLGELIPNVNYIDAAHGMIGMDSNFGSPDIYFEAANASGSIYDVYHGF 398
Qy 278 VYNTATYTLKPEPMTATQOVGQPVESANFYQWNPDIADYADVLYSLVAAGPSSI 337
Db 399 ISNTANYSATVEBYLDVNGTVNGQLIRQSKWMDWNOYDELSPVQDLRLQALIAKGRSDI 458
Qy 338 GMLIDIRLNGMGSPNPGPSTATDVNTFYVNOCKIDLRHGRIGMCONAGLQGPQASP 397
Db 459 GMLIDISRNMGWGPKNRPTGPSSSTDINTYVDESRIIDRIHPGMWCONAGLGEPRPTVP 518
Qy 398 TEPFNAHLDAYVWIKRPGESDGTSAASDPPTGKSDPMCDPTVTSYGVLTN---ALPNS 454
Db 519 A----PGVAYVWVKKRPGESDGAISEIIPNDEGKGFPRMCDPTVQGNARONNPSGALPNA 574
Qy 455 PLAGOFPPAOFDOLVANARP 474
Db 575 PISGHWFSAOFRELLANARP 594

RESULT 5
Q9S3V4 PRELIMINARY; PRT; 351 AA.
ID Q9S3V4;
AC Q9S3V4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 1,4-BETA-CELLULOBIODIOLASE A (EC 3.2.1.91) (FRAGMENT).
GN CELF.
OS Cellulomonas flavigena.
OC Bacteria; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
OX NCBI_Taxid=1711;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDB531;
RA Herrera A., Gutierrez A., Salgado L.M., Ponce-Neoyola T.;
RT "Molecular characterization of cellulases from Cellulomonas
RT flavigena.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBD databases.
DR EMBL; AF17344; AAD48493.1; -.
DR HSSP; P07987; 1CB2.
DR InterPro; IPR001524; Glyco_hydro-6.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR PRINTS; PR00733; GLHTRLAS6.
DR PRODOM; PD00373; GLHTRLAS6.
DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KW hydrolase; Glycosidase.
FT NON_TER 351
FT SEQUENCE 351 AA; 37615 MW; BCDE63B5DF335C2 CRC64;

Query Match 13.1%; Score 722.5; DB 2; Length 351;
Best Local Similarity 48.1%; Pred. No. 4.4e-29;
Matches 166; Conservative 45; Mismatches 105; Indels 29; Gaps 13;

Qy 21 PAISKRLRAGVLAGANVSIASL-VPLAMQ--HPAIAA-THVDNPIYCATFTFVNPYNA-OE 75
Db 6 PSYARRRRRTV--GAFTTALAVALVPLTSLSPASAAEAEKRVDPYAGAVOYVNPNTAASS 63
Qy 76 VOSEANQTNATLAARVRYSTYSTAVMMDRIRAINGVNGPGLTITTYLDAALSOOGGT-T 134
Db 64 VARSAKQSDASIAAKMQVYAKOPTAVMDRISAITGNADGKGLKHLDAVAOQKRAAG 123
Qy 135 PEYIEIYVDLPGRDCAALASNGELPATAAGLQTYETQYIDPIASTILSNPKYSLRIVTI 194
Db 124 PLVFNFLVYNLPGRDCTYALASNGELPATDAGLARQSERIDPIVALLEGREYNLRIAAT 183
Qy 195 IEPDSLPNAVNTNKSIOACATAVPYRYEGGLEXTALTKLHAIPNVYIYMDAAHSGMLGNPNA 254
Db 184 IEPDSLPLNLTNISASCOQAAPIYRQGVKYLADKLIALGNVNYIDIGISGMLGMDSNA 243
Qy 255 SGVYQ---EVQKVLNASTIGVNGIDGFVTNTANTPLKE--PEMTATQOVGQPVESAN-F 308
```

```
Db 244 GPAAKFAEYAKTTNA--GPASTIGFVSDVANTTPLEEPYRPERPSGEGNPIREQYV 301
Qy 309 YQWNPDIADY--YAVDLY--SRLVAAGPSSIGMLIDTLRNGMG 349
Db 302 TSGNNELSTFNRNLRYGATLHRVGSFGRGPRPE-----RRPWG 337

RESULT 6
Q96V97 PRELIMINARY; PRT; 490 AA.
ID Q96V97;
AC Q96V97;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLULOBIODIOLASE II-Like CELLULASE CELI.
GN CELI.
OS Orphiomyces sp. PC-2.
OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
OC Neocallimasticeae; Orphiomyces.
OX NCBI_Taxid=50059;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PC-2;
RX MEDLINE=21405733; PubMed=11514516;
RA Steendakkers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
RA Ljungdahl L.G., Op Den Camp H.J.M.;
RT "Noncatalytic docking domains of cellulosomes of anaerobic fungi.";
RL J. Bacteriol. 183:5325-5333(2001).
DR EMBL; AF177205; AAL01212.1; -.
KW Hydrolase.
SQ SEQUENCE 490 AA; 54051 MW; 85F40D1280759886 CRC64;

Query Match 12.7%; Score 705; DB 3; Length 490;
Best Local Similarity 39.0%; Pred. No. 5e-28;
Matches 167; Conservative 54; Mismatches 147; Indels 60; Gaps 12;

Qy 50 PAIAATHVDNPNVAGATFFVNPYNAOEVOSEANQTNATLAARVRYSTYSTAVMMDRIRIA 109
Db 118 PGEQYTHSGNPFSGVEFFLNPYYAEVDALIEQMTNSLSKAKKEKKYTSNAIMLDTIK- 176
Qy 110 INGVNGSPGLTITLDAALSOOGCTTEBYIRI-VIYDLPGRDCAALASNGELPATAAGLOT 168
Db 177 ---NMHQLWLETMLKGAALQOQFTGSKVILVFVYVYDLPGRDCHALASNGELANDSDAQR 232
Qy 169 YETQYIDPIASILSNPKYSLRIVTIIEPDSLNAVYTNM-SIOACATAVPYRYEGGLEVAL 227
Db 233 YKTEYIDVIEKLT--YKSGPYVLLIVERPSLNLVTLNLDSTPACRDSERYTLDGHAFLI 290
Qy 228 TKLHAIPNVYIYMDAAHSGMLGNPNNASGYVOEVOKVLNASIGVNG-IDGFVTNTANTYP 286
Db 291 QKLGVLPHVAMYLDIGHAFLGMDNREKAGKYKYKI--SSGPKGVNRGFTDNVANVYTP 348
Qy 287 LKEPMTATQOVGQPVESANFYQWNPDIADYAVDLYSRLVAAGPSSIGMLIDTLRN 346
Db 349 WEDEPTLSRGPEI-----EWNPCDEKRYTLELMKDKRAAGIESVYVEVCDTSRN 397
Qy 347 GMGCPNPEPTSPATDVNPFVNOCKIDLRHGRIGMCONAGLQGPQASPTDFPNAHLID 406
Db 398 G-----HKYD-RKHPEKCNQISGVGARGARQASPVSGMD-YLD 433
Qy 407 AYVWIKRPGESDGTSAASDPPTGKSDPMCDPTVTSYGVLTNALPNSPIAGOWFPAQPD 466
Db 434 AFYWIKRPGESDGTSDTS---DTSARVDGYCGH-----DTAMKPAPAEAGQWFOKHFE 480
Qy 467 QIVANARP 474
Db 481 QGLENNAP 488

RESULT 7
Q96V98
```


RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE-97000351; PubMed-8843436;
 RA Redenbach M., Kleiser H.M., Denapate D., Eichner A., Cullum J.,
 RT Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RL the 8 Mb streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL133210; CAB61599.1; .
 DR HSSP; P07986; IEXG.
 DR InterPro; IPR001919; CBD_2.
 DR InterPro; IPR002594; Glyco_hydro_12.
 DR Pfam; PF00553; CBD_2; 1.
 DR Pfam; PF01670; Glyco_hydro_12; 1.
 DR ProDom; PD004316; Glyco_hydro_12; 1.
 DR PROSITE; PS00561; CBD_BACTERIAL; UNKNOWN_1.
 SO SEQUENCE 381 AA; 39199 MW; 97CDBF58679B4EC CRC64;

[illegible]

	RESULT	10
054331		
ID	054331	PRELIMINARY; PRT; 361 AA.
AC	054331;	
DT	01-NOV-1996 (TREMBLrel. 01, Created)	
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	CELLULASE B PRECURSOR.	
CN	CELB.	
OS	Streptomyces lividans.	
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces	
OX	NCBI_taxid=1916;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=66;	
RX	MEDLINE=94288649; PubMed=8017952;	
RA	Wiltman S., Shaeck F., Kluepfel D., Morosoli R.;	
RT	"Purification and Characterization of the CelB endoglucanase from	

RT Streptomyces lividans 66 and DNA sequence of the encoding gene.";
RL Appl. Environ. Microbiol. 60:1701-1703(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-66;
RA Shareck F.;
RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; U04629; AAB71950.1; -.
DR HSSP; P07986; IEXG.
DR InterPro: IPR001919; CBD_2.
DR InterPro: IPR002594; Glyco_hydro_12.
DR pfam; PF00553; CBD_2; 1.
DR pfam; PF01670; Glyco_hydro_12; 1.
DR ProDom; PD004316; Glyco_hydro_12; 1.
DR PROSITE; PS00561; CBD_BACTERIAL; UNKNOWN_1.
KW Signal.
FT SIGNAL
FT CHAIN 1 40 POTENTIAL.
FT CHAIN 41 381 CELLULASE B.
SQ SEQUENCE 361 AA; 39239 MW; A7E9BFE550FA24EC CRC64;

	Query Match	11.7%	Score 646.5	DB 2	Length 381	
	Best Local Similarity	40.7%	Pred. No. 3.2e-25			
	Matches 155	Conservative 45	Mismatches 144	Indels 37	Gaps 14	
QY	684	GLAAG----	TAATYTVAAVDAAGNTASAPSTVEDCTPCPGNNGVTSVODGEYRVQTNEMWN	738		
DB	13	GLLALGAVLAFALVLSLVTAAPAAQDPT--ICEP----	EGTITIQ-GRVYQNNRRG	65		
QY	739	SSAQCCLINTATGTAMVYSTANFSGGTGAPATYTSYKGCWMCCTTKNMGMPQISOI	798			
DB	66	SNAPCCVNA-IDTG-FRTQADGASAPTNAGAPKSYSPYSEFNGCHYTC-N	SPGTDLPRLDLIV	122		
QY	799	GSAAVSMSTQVYSSGADYADYDITMNSPTTGTGPNGTREIMTLWSRGVOPFGSOTATG	858			
DB	123	SNAPSPSISVGFDAVAVNASTYDITMDLPKRTIG-VNQETITMTEFNRYGFIQYIGSPVGT-	180			
QY	859	VTVAGHTMNVWVGQGTSMKLIISVLTLPGATSIISNLDLKAIFDAARGLSNTSDYLDVE	918			
DB	181	ASVGRTRTEWVMGGSGNSNDVLSFV-APSAISGMSFVDVDFRATVARGLAENDMTLTSVQ	239			
QY	919	AGFELIMOGGOGIGSNSFSVSY---TSGGV-----	ACRATYVNVNSDWSGCFATYV	966		
DB	240	AGFEFWQAGALVAVSFSSTVETGTPGGIDPDPCGSPSCAVSYGTNV-WQDGFADYV	298			
QY	967	TNTGSRATNGMTVAANSFGNGQTVTYMYNTALQSGASVYATNLVYENNVYIQPQSTTFGEN	1026			
DB	299	TNTGTAPYDQMGVLATLPDGCGRITTNAMNNLSLPPSSGVSATATASHNARATAPGSLSPGRQ	358			
QY	1027	GSYSGTNAAPT-----LSC	T 1041			
DB	359	GTYGGAFAEPETGFRLNGTACT	379			

RESULT	11		
09K1H1			
ID	09K1H1	PRELIMINARY:	PRT: 371 AA.
AC	09K1H1:		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DI	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DR	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	CELLULASE 12A.		
GN	CEL12A.		
OS	Streptomyces sp. 11A68.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
OX	NCBI_taxid=133452;		
RM	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-11A68:		
RA	van Solingen P., Meijer D., van der Kleij W.A.H., Barnett C.C.,		
RA	Bolle R., Power S.D., Jones B.E.;		
RT	"Cloning and expression of an endocellulase gene from a novel		

RT Streptomycete isolated from an East African soda lake.
 RL Submitted (Feb-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE233376; AAF91283.1; -
 DR HSSP: P07986; 1EXG.
 DR InterPro: IPR002594; Glyco_hydro_12.
 DR InterPro: IPR001230; Prenylth.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR ProDom: PD004316; Glyco_hydro_12; 1.
 DR PROSITE: PS00294; PRENYLATION; UNKNOWN_1.
 SQ SEQUENCE 371 AA; 38481 MW; 0E1BC4288A148914 CRC64;

Query Match 11.4%; Score 633.5; DB 2; Length 371;
 Best Local Similarity 40.0%; Pred. No. 1.4e-24;
 Matches 148; Conservative 52; Mismatches 131; Indels 39; Gaps 15;

QY 695 VAAVDAAGNTSAPSTPDCPTGPNQ-----GVTSVODGEYRQVTNEMNSAOCCLINT 749
 DB 16 LASLGLALTAAPAO-----ANQOICDRYGTITD-RVYVONNMWGTSAQC--INV 65
 QY 750 ATGAMVTSTANSGGTGAPATYPSIYKCHMGNCCTKNVGMPIQISGSAVTSMTQ 809
 DB 66 TNGFETIQADSGVPTNGAKPSYVDGCHYGKCAPRT-LPRKISSIGSAPSSTYR 124
 QY 810 VSSGADVADIVMTNSTPTTGTGPNCTEIMILWNSRGVOPFGSQATGVTVAGHTWNV 869
 DB 125 TNGGVYNAADIVMLDPTPTNG-VNRTIEMIFNRVGPVQIGSPVGT-AHVGGRSWEV 182
 QY 870 QGGQTSKIIISVLTPTGATISNLDKAFADAARGLMTSD-YLIDVGEFINGGQ 928
 DB 183 TGSNGSDVLSF-LAPSAISSMSPVDK-FVDQAVSHGLATPDMYLTLSIQAGFPMWGGT 240
 QY 929 GLSNSNSFSVTSVG-----VACRATVYVNSDMGSGFTATVYTNNGSRATNGTV 979
 DB 241 GLAVNSFSVAVNMGNGGPGTGPACQVSTYHT-WPGFTVDTITITNGSTVDGHEL 299
 QY 980 AMFPGNQVYTNMTALTQSGASVATNLXSN-NVIOPOSTGTFNGSYGTN-AAPT 1037
 DB 300 DFTLPAGHTVTSWNALISPAAGAVTARSTGNGRIANAGTSGFGTSSGAGTAPA 359
 QY 1038 -----LSCPT 1041
 DB 360 GARLNGTSCPT 369

RESULT 12
 059963 PRELIMINARY; PRT; 382 AA.
 AC 059963;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CELLULASE (EC 3.2.1.4).
 GN BGLS.
 OS Streptomyces rochei (Streptomyces parvullus).
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_Taxid=1928;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A2;
 RX MEDLINE=95011642; PubMed=7523249;
 RA Perito B., Hanhart E., Irdant T., Iqbal M., McCarthy A.J.,
 RA Mastromel G.;
 RT "Characterization and sequence analysis of a Streptomyces rochei A2";
 RL Gene 148:119-124(1994).
 DR EMBL: X73953; CAAS2139.1; -
 DR HSSP: P07986; 1EXG.
 DR InterPro: IPR001919; CBD_2.
 DR InterPro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR ProDom: PD004316; Glyco_hydro_12; 1.

DR PROSITE: PS00561; CBD_BACTERIAL; UNKNOWN_1.
 KW Hydrolase; Glycosidase.
 SQ SEQUENCE 382 AA; 39398 MW; 21C014342EFC6565 CRC64;

Query Match 11.4%; Score 629.5; DB 2; Length 382;
 Best Local Similarity 39.9%; Pred. No. 2.3e-24;
 Matches 147; Conservative 48; Mismatches 148; Indels 25; Gaps 9;

QY 685 LAAGTATVYVAAVDAAGNTSAPSTPDCPTGPNQNGVTSVODGEYRQVTNEMNSAOC 744
 DB 11 LRAVSAALTAALALATAPADDTT-CEPFGSTVIO-GRVYVONNMWGTSAQC 68
 QY 745 LTIMATGAMVTSTANSGGTGAPATYPSIYKCHMGNCCTKNVGMPIQISGSAVTS 804
 DB 69 VT--ATDSGFRVTDADSGVPTNGAKPSYVFNCHYINC-SPTALPARISGSSAPSS 125
 QY 805 WSTQVSSGADVADIVMTNSTPTTGTGPNCTEIMILWNSRGVOPFGSQATGVTVAGH 864
 DB 126 ISYGFVNNAVYNAASYDMLDPTPTDQ-VNRTIEMIFNRVGOIQIGSQVGT-ASVAGR 183
 QY 865 TWNVMOGQTSKIIISVLTPTGATISNLDKAFADAARGLMTSDYLLDVENGEFIW 924
 DB 184 TWEVWSGNGINDVLSFV-APSAIISMSFVMDVFRATVAGLAGNDVYLTLSIQAGEPW 242
 QY 925 QGGQGLSNSFSVTSVG-----VACRATVYVNSDMGSGFTATVYTN 968
 DB 243 QNGAGLAVNSFSSTVYVNGSGNPGDPNGDPGTPACTVSYATNV-WPGFTANVTYN 301
 QY 969 TGRATNGVTVAMSGNQVYTNMTALTQSGASVATNLXSN-NVIOPOSTGTFNGS 1028
 DB 302 NGSAPVDSGWRALFPLPSGQSVVHAMNASVSPSCAVTATCPAESARIAAGSOSFGFOGA 361
 QY 1029 YSGTNAAP 1036
 DB 362 YSGSFAQP 369

RESULT 13
 002321 PRELIMINARY; PRT; 460 AA.
 AC 002321;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE EXOCELLULOBIOMHYDROLASE.
 GN CBH1T.
 OS Phanerochaete chrysosporium.
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Aphyllophorales; Corticiaceae; Phanerochaete.
 OX NCBI_Taxid=5306;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95110131; PubMed=7811079;
 RA Templears C.A., Birch P.R., Sims P.F., Broda P.;
 RT "Isolation, characterization, and analysis of the expression of the
 RT cbh1 gene of Phanerochaete chrysosporium";
 RL Appl. Environ. Microbiol. 60:4387-4393(1994).
 DR EMBL: S76141; AAB32942.1; -
 DR HSSP: P00725; ZCBH.
 DR InterPro: IPR000254; CBD_fungal.
 DR InterPro: IPR001524; Glyco_hydro_6.
 DR Pfam: PF00734; CBD_1; 1.
 DR Pfam: PF01341; Glyco_hydro_6; 1.
 DR PRINTS: PR00733; GLHDBLASE6.
 DR ProDom: PD001821; CBD_fungal; 1.
 DR ProDom: PD003733; Glyco_hydro_6; 1.
 DR SMART: SM00236; fcbd; 1.
 DR PROSITE: PS00562; CBD_FUNGAL; 1.
 DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
 DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
 KW Hydrolase.
 SQ SEQUENCE 460 AA; 48444 MW; A8BABA40615FE56B CRC64;

```

Query Match      11.2%; Score 618.5; DB 3; Length 460;
Best Local Similarity 36.8%; Pred. No. 1e-23;
Matches 154; Conservative 55; Mismatches 144; Indels 65; Gaps 12;

OY 58 DNPVAGATFEVNPYMAOEVOSEANOTNATLAAMRVSTYSYAVVMMDRITAINGVNGPG 117
DB 105 NNPWTFQJFLSPRYANETAAAKQITDPTLSKASAVNIPFTWLDVAKI-----P 158
OY 118 GLTYLDAALSOOQGT-TPEVIEIYVDLPGRCALASNGELPATVAGLOTETGYIDP 176
DB 159 DGTGYLASALGKSTGTQVQIVYIDLPDRDCAKAKANGESIANNGQANE-NYIQD 217
OY 177 IASISLNPYSSLRITYITTEPDSLPNAVNMSTQACATVPPYEGIEYALTKLHAIPNV 236
DB 218 IYAQID--QFPVRYVAVEIPDSLANVTNLNVQCANAKTYTACVNALFTMLAKV-GV 274
OY 237 YIYMDAASHGMLGMPNNAAGVYQVOKVLNASIGVNGIDGFTVNTANTYPLKEPMTATQ 296
DB 275 YMYMDAGHAGMLGMPANLSPAAQLFTQVWONNKGSPFINGLATVNAVY-----NALQ 326
OY 297 QVGGQPVESANFYQWNPDIDEADYAVDLXSRUYAGFPSSIGMLIDTLRNGMGPNPEPTG 356
DB 327 AASPDPT-----QGNPNYDEIHYNALAPLQAGNMDAT--FTYDQGRSG----- 370
OY 357 PSTATDVNTFVWNSKIDLRQHRGLMCNONGAGIGQPPQASPTDFPNAHLDAVYIKPGE 416
DB 371 -----NINIGLMLGMLGMPNNAAGVYQVOKVLNASIGVNGIDGFTVNTANTYPLKEPMTATQ 413
OY 417 SDGTSASDPPTGKSDPMCDPTTYSYGLVTNALPNSPIAGOWFPAQFDOLVANARP 474
DB 414 CQGTSMSSP-----RYDSTCS-----LPDAQPAPEAGTWFQAYFETLVSAANP 458

RESULT 14
O96VU2 PRELIMINARY; PRT; 444 AA.
AC O96VU2:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELULOSE CEL6B.
GN CEL6B.
OS Lentinula edodes (Shiitake mushroom) (Lentinus edodes).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Tricholomataceae; Lentinula.
OX NCBI_TaxID=5353;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-STAMETS CS-2;
RA Lee C.C., Wong D.W.S., Robertson G.H.;
RT "Cloning and characterization of two cellulase genes from Lentinula
RT edodes."
RT Submitted (AUG-2001) to the EMBL/GenBank/DDbj databases.
DR EMBL: AF411251; AAK95564.1;
DR EMBL: AF411251; AAK95564.1; A08BDD0835FE47C5 CRC64;
SQ SEQUENCE 444 AA; 46368 MW; A08BDD0835FE47C5 CRC64;

Query Match      10.8%; Score 597.5; DB 3; Length 444;
Best Local Similarity 35.3%; Pred. No. 1.1e-22;
Matches 147; Conservative 60; Mismatches 143; Indels 67; Gaps 14;

OY 59 NPYAGATFEVNPYMAOEVOSEANOTNATLAAMRVSTYSYAVVMMDRITAINGVNGPG 118
DB 92 NPTGEIYILSPYANETAAAVTQISDPTTAAAKAVANIPFTIMDQYAKV-----PD 145
OY 119 LTTYL-DALASOQGTTPVIEIYVDLPGRCALASNGELPATVAGLOTETGYIDP 177
DB 146 LGTYLDAALSOOQGTTPVIEIYVDLPGRCALASNGELPATVAGLOTETGYIDP 204
OY 178 ASILSNPKYSSLRITYITTEPDSLPNAVNMSTQACATVPPYEGIEYALTKLHAIPNV 237
DB 218 IYAQID--QFPVRYVAVEIPDSLANVTNLNVQCANAKTYTACVNALFTMLAKV-GV 274

```

```

DB 205 VAOIK--QYPDVHVAVIEPDSLANVTNLNVQCANAKTYTACVNALFTMLAKV-GV 261
OY 238 IYMDAASHGMLGMPNNAAGVYQVOKVLNASIGVNGIDGFTVNTANTYPLKEPMTATQ 297
DB 262 MYMDAGHAGMLGMPANLSPAAQLFTQVWONNKGSPFINGLATVNAVY-----NALQ 326
OY 298 VGGQPVESANFYQWNPDIDEADYAVDLXSRUYAGFPSSIGMLIDTLRNGMGPNPEPTG 356
DB 317 ---DPTT-----QGNPNYDEIHYNALAPLQAGNMDAT--FTYDQGRSG----- 370
OY 358 STATDVNTFVWNSKIDLRQHRGLMCNONGAGIGQPPQASPTDFPNAHLDAVYIKPGE 416
DB 355 -----VQIRQWMDMCMVNLGAGETOP--TTNIGSSLDISYVWAKGEGC 398
OY 418 DGTSAASDPPTGKSDPMCDPTTYSYGLVTNALPNSPIAGOWFPAQFDOLVANARP 474
DB 399 DGTSMSSP-----RYDAHCG-----LPDAQPAPEAGTWFQAYFETLVSAANP 458

RESULT 15
O9CIR4 PRELIMINARY; PRT; 443 AA.
AC O9CIR4:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLOBIOHYDROLASE.
GN CBHI1-1.
OS Lentinula edodes (Shiitake mushroom) (Lentinus edodes).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Tricholomataceae; Lentinula.
OX NCBI_TaxID=5353;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-L54;
RA Zhao J., Deane E.E., Kwan H.S.;
RT "Molecular cloning, characterization, and differential expression of a
RT cellobiohydrolase gene from the edible mushroom Lentinula edodes."
RT Submitted (MAR-2000) to the EMBL/GenBank/DDbj databases.
DR EMBL: AF244369; AAK28357.1;
DR HSSP: P00725; IAZ6.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR001524; Glyco_hydro_6.
DR Pfam: PF00734; CBD_1; 1.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PR00733; GLHYDRLASE6.
DR Prodom: PD001821; CBD_fungal; 1.
DR Prodom: PD003733; Glyco_hydro_6; 1.
DR SMART: SM00236; fcbd.1.
DR PROSITE: PS00362; CBD_FUNGAL; 1.
DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KW Hydrolase.
SQ SEQUENCE 443 AA; 46290 MW; EA9B5788F79EC25A CRC64;

Query Match      10.6%; Score 589.5; DB 3; Length 443;
Best Local Similarity 33.4%; Pred. No. 2.9e-22;
Matches 158; Conservative 65; Mismatches 173; Indels 77; Gaps 16;

OY 7 SGNCRYQKGTMPAIRAGVLAGAVSTIASIVPLAMQHPATIAATVNDPVGATP 66
DB 41 SGATC-----TVVNAVYSOCLPGSAPPTSSIGTGTTSASAGSTGTPPAAGNPF 94
OY 67 ----FVNPYMAOEVOSEANOTNATLAAMRVSTYSYAVVMMDRITAINGVNGPG 122
DB 95 TEOITYLSPYANETAAAVTQISDPTTAAAKAVANIPFTIMDQYAKV-----PD 145
OY 123 L-DALASOQGTTPVIEIYVDLPGRCALASNGELPATVAGLOTETGYIDP 181
DB 149 LADASAKOKSEKKNYLVOIVYVDLPDRDCAALASNGEFTIADNGEANYH-DYIDQVAGI 207
OY 182 SNPKYSSLRITYITTEPDSLPNAVNMSTQACATVPPYEGIEYALTKLHAIPNV 241

```

```

Db      208  K--QYPDVHVAVIIEPPDSLANTLVNLSVAKCANAQTTTECVTYAMQQLSAV--GVTMYILD 264
OY      242  AAHSGWLGMPNNASGYVOEYOKVLNASIGVNGIDGFTVTNTANTTPLKEPMTATQOYGQ 301
Db      265  AGHAGWLGMPNNASGYVOEYOKVLNASIGVNGIDGFTVTNTANTTPLKEPMTATQOYGQ 316
OY      302  PVESANFYQWNPDIADYAVDLYSRLVAGFPSSIGMLIDTLRNGWGGPNEPTGPSTAT 361
Db      317  PIT-----OGDPNYDEMILYIEALAPLL--GSFPAH--FIVDQGRSG----- 353
OY      362  DVNTFVNQSKIDKROHRLGMLCNGAGLGQPPQASPTDFPNAHLDAYVWIKPGESDGT 421
Db      354  -----VQDIRQWGMCMCNVLGAGFTQP--TTNIGSSLIDSIWVWPGGECDDGT 401
OY      422  AASDPTTGKSDPMCDPTTYSYGVLTNALPNSPIAGOMFPPOEDOLVANARP 474
Db      402  NTSSP-----RYDAHCG-----LPDATPNAPEAGTWFQATFETLIVEKANP 441

```

Search completed: August 29, 2002, 16:37:51
 Job time: 961 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:22:35 ; Search time 47.58 Seconds
(without alignments)
535.433 Million cell updates/sec

Title: US-09-917-384-6
Perfect score: 5536
Sequence: 1 MERTOGSGRNCRYGRTTRM.....GFGNGSYSGTNAAPTLCSTAS 1043

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5a_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5b_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5a_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5b_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/5a_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/5b_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	646.5	11.7	381	4	US-09-216-295-22 Sequence 22, Appl
2	639.5	11.6	352	4	US-09-286-691-26 Sequence 26, Appl
3	639.5	11.6	352	4	US-09-687-147-26 Sequence 26, Appl
4	637	11.5	386	4	US-09-321-981-5 Sequence 5, Appl
5	633.5	11.4	371	4	US-09-104-308-1 Sequence 1, Appl
6	633.5	11.4	371	4	US-09-321-981-1 Sequence 1, Appl
7	564.5	10.2	312	4	US-09-216-295-21 Sequence 9, Appl
8	556.5	10.1	461	3	US-09-118-319-9 Sequence 25, Appl
9	555	10.0	360	4	US-09-286-691-25 Sequence 25, Appl
10	555	10.0	360	4	US-09-687-147-25 Sequence 25, Appl
11	548.5	9.9	470	3	US-09-118-319-8 Sequence 8, Appl
12	544	9.8	457	3	US-09-142-759-1 Sequence 1, Appl
13	538	9.7	360	4	US-09-286-691-24 Sequence 24, Appl
14	538	9.7	360	4	US-09-687-147-24 Sequence 24, Appl
15	538	9.7	360	4	US-09-687-147-24 Sequence 12, Appl
16	538	9.7	360	4	US-09-687-147-24 Sequence 12, Appl
17	538	9.7	360	4	US-09-687-147-24 Sequence 12, Appl
18	467	8.4	1751	4	US-08-382-4520-12 Sequence 43, Appl
19	449	8.1	1426	4	US-09-136-574A-43 Sequence 4, Appl
20	444.5	8.0	449	4	US-09-118-319-7 Sequence 7, Appl
21	444.5	8.0	449	4	US-09-286-691-4 Sequence 4, Appl
22	444.5	8.0	449	4	US-09-687-147-4 Sequence 4, Appl
23	436	7.9	428	3	US-09-118-319-5 Sequence 23, Appl
24	434.5	7.8	326	4	US-09-286-691-23 Sequence 23, Appl
25	434.5	7.8	326	4	US-09-687-147-23 Sequence 23, Appl
26	431.5	7.8	432	3	US-09-118-319-2 Sequence 2, Appl
27	417.5	7.5	551	2	US-09-033-537A-1 Sequence 1, Appl

28	412	7.4	459	3	US-09-118-319-6 Sequence 6, Appl
29	412	7.4	459	4	US-09-286-691-2 Sequence 2, Appl
30	412	7.4	459	4	US-09-687-147-2 Sequence 2, Appl
31	407.5	7.4	260	4	US-09-216-295-23 Sequence 23, Appl
32	406	7.3	521	1	US-08-276-213-3 Sequence 3, Appl
33	388.5	7.0	700	2	US-07-862-588B-2 Sequence 2, Appl
34	376	6.8	616	4	US-09-136-574A-47 Sequence 47, Appl
35	371.5	6.7	493	4	US-09-198-955A-12 Sequence 12, Appl
36	371.5	6.7	493	4	US-09-198-955A-12 Sequence 9, Appl
37	362	6.5	167	5	PCR-US95-13813-9 Sequence 16, Appl
38	269.5	4.9	348	4	US-09-216-295-16 Sequence 5, Appl
39	260.5	4.7	259	4	US-09-216-295-5 Sequence 2, Appl
40	244.5	4.4	108	2	US-08-585-585A-2 Sequence 2, Appl
41	244.5	4.4	108	2	US-08-249-037C-2 Sequence 2, Appl
42	244.5	4.4	108	2	US-08-788-622B-2 Sequence 2, Appl
43	244.5	4.4	108	3	US-08-788-621B-2 Sequence 2, Appl
44	242.5	4.4	291	4	US-09-286-691-28 Sequence 28, Appl
45	242.5	4.4	291	4	US-09-687-147-28 Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-09-216-295-22
; Sequence 22, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; TITLE OF INVENTION: No. 6268328el Variant EgIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Streptomyces lividans Celb
; US-09-216-295-22

Query Match 11.7%; Score 646.5; DB 4; Length 381;
Best Local Similarity 40.7%; Pred. No. 3.9e-38;
Matches 155; Conservative 45; Mismatches 144; Indels 37; Gaps 14;

QY	684	GLAAG-----TAYTYVAAYDAAGNTSADSTPYDCTPGFNONGYTSYODGEYRVQNTENM	738
DB	13	GLAALAGVALAFALVSLVTAAAPADDT--ICEP---FETTYIO-GRIVQNNRMG	65
QY	739	SSAOCCLTIMTACAWTVSTANFSGGTGAPATYPSIYKCHMGNCYTRNVGMP10ISOI	798
DB	66	STAPQCYTA-TDGG-FEYTOADGSAPTNKAPKSPYFVFNCHTNC-SPTGDLPAVLDTIV	122
QY	799	GSAYTSMSTQVSSGADVAYDIWTNSTPTTQPNCTEIMWLNSRGVQPPGSGTATG	858
DB	123	SAAPSSTSYGVDDAVYANASYDIWLDPTARTDG-VNQTETIMFNRRVGPLOPGSPVGT	180
QY	859	VTVAGHTVNWVQSGOTSMKIISYLVLTGATGISISLKDALFADAAAGSINTSDYLDVE	918
DB	181	ASVGRTVEWVSGNGSNDVLEFV-APSAISGMSFVDMYDVRATVANGLAENMWLTSYVQ	239
QY	919	AGFEIMOGGGLGNSFESVSFV---TSGCV-----ACRATVNSDMGSGFATVTV	966
DB	240	AGFEWQMGAGLAVNSFESVETGTPDGDGCPSCACAVSYGTVV-WQDGFADTV	298
QY	967	TNTGSRATNGTVANFSGNCTVNYMTALTOSGASVTATNLNSNNVIOGOSTTFGFN	1026
DB	299	TNTGTAPYDQWQLATLPSGQRTINAMNASLPSGSGVTATGASHNARIAPGSLSFQ	358
QY	1027	GSYSGTNAAPT-----LSCF 1041	

Db 359 GTYGGAFAEPTGRLNCTACT 379

RESULT 2

US-09-286-691-26

Sequence 26, Application US/09286691

Patent No. 6190189

GENERAL INFORMATION:

APPLICANT: Li, Xin-Liang

APPLICANT: Ljungdahl, Lars G.

TITLE OF INVENTION: Cellulases and Coding Sequences

FILE REFERENCE: 42-96

CURRENT APPLICATION NUMBER: US/09/286,691

CURRENT FILING DATE: 1999-04-05

EARLIER APPLICATION NUMBER: US 60/027,883

EARLIER FILING DATE: 1996-10-04

EARLIER APPLICATION NUMBER: PCT US97/18008

EARLIER FILING DATE: 1997-10-03

NUMBER OF SEQ ID NOS: 29

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 26

LENGTH: 352

TYPE: PRT

ORGANISM: Agaricus bisporus

US-09-286-691-26

Query Match 11.6%, Score 639.5; DB 4; Length 352;

Best Local Similarity 36.6%, Pred. No. 1.1e-37;

Matches 153; Conservative 62; Mismatches 134; Indels 69; Gaps 12;

QY 59 NPYAGATFFVNPYMAOEYSEANOTNATLAKMRVSTYSTAVWMDRIALINGVNGPG 118

Db 4 NPYGKTWVLSPEYADEVAQAADISNPSLATKASVAKIPTFVWFVDAKV-----PD 57

QY 119 LTTYDAALSOOGCTTPVEIVYIDLPGRCALASNGELPATNAGIQTETQYIDPIA 178

Db 58 LGGYIADARSKNQ-----LVQIVYIDLPGRCALASNGEFLSLANDGLNKTK-NIVDQIA 111

QY 179 SILSNPKYSLRIYVITIEPDSIIPNAVNMISIOACATAPVYEOGIEVALTKLHAIPNYI 238

Db 112 AQIK--QPPDVSVAVIEPDSLNLVTLNVOKCANASAKKEGYIYAVQKLNAY-GVTM 168

QY 239 YMDAASHGWLGMPPNASSGYOVQKVLNASIGVNCIDGFVTNTANTYPLKEPFMTAQOV 298

Db 169 YIDAGHAGWLGMPPANLSPAQOLFQIYRDAGSPRMLRGIAINVANFNALRAS----- 220

QY 299 GGQPVESANFYQWNPDIADAYAVDLSRLVAAFPSSIGMLIDTLRGMWGGPNEPTGPS 358

Db 221 SPDPIT-----QGSNRYDEIRHIEALAPMLSNAGFPAN--FIVDGRSG----- 262

QY 359 TATDVNTEFNOSKIDLROHRLGIMCNGAGLGQPPQASPTDFPNALHDAYWIKPPGESD 418

Db 263 -----VQNIHQMGDMCMVKAGFGQRP---TTMGSLLDAIYVWVRKGEGCD 307

QY 419 GTSASDPTTGKKSPKMDPTTYSYGVLTNALPNSPIAGOMFPQFOQLYANAPAY 476

Db 308 GTSDNSSP-----RFDSHCS-----LSDAHQPAPEAGTWFOAYFETLVANANPAL 352

RESULT 3

US-09-687-147-26

Sequence 26, Application US/09687147

Patent No. 6268198

GENERAL INFORMATION:

APPLICANT: Li, Xin-Liang

APPLICANT: Ljungdahl, Lars G.

TITLE OF INVENTION: Cellulases and Coding Sequences

FILE REFERENCE: 42-96a

CURRENT APPLICATION NUMBER: US/09/687,147

CURRENT FILING DATE: 2000-10-12

PRIOR APPLICATION NUMBER: US 60/027,883

PRIOR FILING DATE: 1996-10-04

PRIOR APPLICATION NUMBER: PCT US97/18008

PRIOR FILING DATE: 1997-10-03

PRIOR APPLICATION NUMBER: 09/286,691

PRIOR FILING DATE: 1999-04-05

NUMBER OF SEQ ID NOS: 29

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 26

LENGTH: 352

TYPE: PRT

ORGANISM: Agaricus bisporus

US-09-687-147-26

Query Match 11.6%, Score 639.5; DB 4; Length 352;

Best Local Similarity 36.6%, Pred. No. 1.1e-37;

Matches 153; Conservative 62; Mismatches 134; Indels 69; Gaps 12;

QY 59 NPYAGATFFVNPYMAOEYSEANOTNATLAKMRVSTYSTAVWMDRIALINGVNGPG 118

Db 4 NPYGKTWVLSPEYADEVAQAADISNPSLATKASVAKIPTFVWFVDAKV-----PD 57

QY 119 LTTYDAALSOOGCTTPVEIVYIDLPGRCALASNGELPATNAGIQTETQYIDPIA 178

Db 58 LGGYIADARSKNQ-----LVQIVYIDLPGRCALASNGEFLSLANDGLNKTK-NIVDQIA 111

QY 179 SILSNPKYSLRIYVITIEPDSIIPNAVNMISIOACATAPVYEOGIEVALTKLHAIPNYI 238

Db 112 AQIK--QPPDVSVAVIEPDSLNLVTLNVOKCANASAKKEGYIYAVQKLNAY-GVTM 168

QY 239 YMDAASHGWLGMPPNASSGYOVQKVLNASIGVNCIDGFVTNTANTYPLKEPFMTAQOV 298

Db 169 YIDAGHAGWLGMPPANLSPAQOLFQIYRDAGSPRMLRGIAINVANFNALRAS----- 220

QY 299 GGQPVESANFYQWNPDIADAYAVDLSRLVAAFPSSIGMLIDTLRGMWGGPNEPTGPS 358

Db 221 SPDPIT-----QGSNRYDEIRHIEALAPMLSNAGFPAN--FIVDGRSG----- 262

QY 359 TATDVNTEFNOSKIDLROHRLGIMCNGAGLGQPPQASPTDFPNALHDAYWIKPPGESD 418

Db 263 -----VQNIHQMGDMCMVKAGFGQRP---TTMGSLLDAIYVWVRKGEGCD 307

QY 419 GTSASDPTTGKKSPKMDPTTYSYGVLTNALPNSPIAGOMFPQFOQLYANAPAY 476

Db 308 GTSDNSSP-----RFDSHCS-----LSDAHQPAPEAGTWFOAYFETLVANANPAL 352

RESULT 4

US-09-321-981-5

Sequence 5, Application US/09321981

Patent No. 6287839

GENERAL INFORMATION:

APPLICANT: Genencor International, Inc.

TITLE OF INVENTION: No. 6287839el Cellulase Producing Actinomycetes,

TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same

FILE REFERENCE: GC540-2

CURRENT APPLICATION NUMBER: US/09/321,981

CURRENT FILING DATE: 1999-05-28

PRIOR APPLICATION NUMBER: PCT/US99/11971

PRIOR FILING DATE: 1999-05-28

PRIOR APPLICATION NUMBER: 09/104,308

PRIOR FILING DATE: 1998-06-24

PRIOR APPLICATION NUMBER: 08/974,042

PRIOR FILING DATE: 1997-11-19

NUMBER OF SEQ ID NOS: 5

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 5

LENGTH: 386

TYPE: PRT

ORGANISM: Unknown

FEATURE:

OTHER INFORMATION: Nearest "neighbor" - Streptomyces

OTHER INFORMATION: thermobaceous
US-09-321-981-5

Query Match 11.5%; Score 637; DB 4; Length 386;
Best Local Similarity 40.6%; Pred. No. 1.9e-37;
Matches 149; Conservative 48; Mismatches 134; Indels 36; Gaps 15;

QY 698 VDAAGTASAPSTPYDCTPGPNON-----GTVSVODGEYRVQTNEMNSSAOCLTINTATG 752
DB 31 VSTVIGTALAGPPAQA-----NOQICDRYGTITID-RVYVONNRNGTSATQC--INVGN 83
QY 753 AMVTSTANFSGTGAPATYPSYKGCWNGCTTKNMGPIQISQISAVTSKSTQVSS 812
DB 84 GFETIADGVSPTNGAPKSPSYVDGCHGNCAPRTT-LPMRISISGASPSVSRYXTGN 142
QY 813 GAYDAVDITNSTPTTGTGPNGTETIMILNSRGVQPGFSQATGVTVAAGHTWVWQO 872
DB 143 GYVNAADIMIDPTPTPTNG-VNRTEIMTFNRRVGPVOPISPVGT-AHYGGRSMEVWTS 200
QY 873 QTSKRTISYVLTGATSTISNLDKAIADAAAGSLNTSD-YLLDVEAGEFIWQGGGLG 931
DB 201 NGSNDVLSF-LAPSAISWSFVDVD-FYDQAVSHGLATPDWYLTLSIQAGEFPEWEGGTGLA 258
QY 932 SNSFSVSVTSGG-----VACRATYVNSDMGSGFTATVTNTGSRATNGMTVAMS 982
DB 259 VNSFSANVAGGNGTGPCTPAACOVSYSTHT-WPGGFTVDTTINTGSPVDGEMELDT 317
QY 983 FGNGOTVNTYNTALTQSGASVATNLYSN-NVIOPGOSTTGFNGSISGNT-AAPT--- 1037
DB 318 LPAGHTVTSANNALISPAAGAVTARSTGNSGRILANGTOSFGOGTSSTGTFNAPAGGR 377
QY 1038 ---LSC 1041
DB 378 LMGTSCT 384

RESULT 5
US-09-104-308-1
Sequence 1, Application US/09104308
Patent No. 6187577
GENERAL INFORMATION:
APPLICANT: Jones, Brian E.
APPLICANT: Van Der Kleij, Wilhelmus A.H.
APPLICANT: Weyler, Walter
TITLE OF INVENTION: No. 6187577el Cellulase Producing Actinomycetes,
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: Genecor International, Inc.
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/104,308
FILING DATE: 24-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/974,042
FILING DATE: 19-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Stone, Christopher L.
REGISTRATION NUMBER: 35,696
REFERENCE/DOCKET NUMBER: GC539
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-846-7555

TELEFAX: 650-845-6504
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 371 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-104-308-1

Query Match 11.4%; Score 633.5; DB 4; Length 371;
Best Local Similarity 40.0%; Pred. No. 3.1e-37;
Matches 148; Conservative 52; Mismatches 131; Indels 39; Gaps 15;

QY 695 VAADAAGTASAPSTPYDCTPGPNON-----GTVSVODGEYRVQTNEMNSSAOCLTINT 749
DB 16 LASIGALLTAAAPAO-----ANQICDRYGTITID-RVYVONNRNGTSATQC--INV 65
QY 750 ATGAMTSTANFSGTGAPATYPSYKGCWNGCTTKNMGPIQISQISAVTSKSTQ 809
DB 66 TGNGETIADGVSPTNGAPKSPSYVDGCHGNCAPRTT-LPMRISISGASPSVSRY 124
QY 810 VSSGAYDAVDITNSTPTTGTGPNGTETIMILNSRGVQPGFSQATGVTVAAGHTWV 869
DB 125 TGNVNAADIMIDPTPTPTNG-VNRTEIMTFNRRVGPVOPISPVGT-AHYGGRSMEV 182
QY 870 QGQOTSMKRTISYVLTGATSTISNLDKAIADAAAGSLNTSD-YLLDVEAGEFIWQGG 928
DB 183 TGSNSNDVLSF-LAPSAISWSFVDVD-FYDQAVSHGLATPDWYLTLSIQAGEFPEWEG 240
QY 929 GLGNSFSVSVTSGG-----VACRATYVNSDMGSGFTATVTNTGSRATNGMTV 979
DB 241 GLAVNFSAVNAGGNGTGPCTPAACOVSYSTHT-WPGGFTVDTTINTGSPVDGEMEL 299
QY 980 AMSFGNGOTVNTYNTALTQSGASVATNLYSN-NVIOPGOSTTGFNGSISGNT-AAPT 1037
DB 300 DFTLPAGHTVTSVNNALISPAAGAVTARSTGNSGRILANGTOSFGOGTSSTGTFNAP 359
QY 1038 ---LSC 1041
DB 360 GARLNGTCTP 369

RESULT 6
US-09-321-981-1
Sequence 1, Application US/09321981
Patent No. 6287839
GENERAL INFORMATION:
APPLICANT: Genecor International, Inc.
TITLE OF INVENTION: No. 6287839el Cellulase Producing Actinomycetes,
NUMBER OF SEQUENCES: 2
FILE REFERENCE: GC540-2
CURRENT APPLICATION NUMBER: US/09/321,981
CURRENT FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: PCT/US99/11971
PRIOR FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: 09/104,308
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 08/974,042
PRIOR FILING DATE: 1997-11-19
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FASTSEQ for windows Version 3.0
SEQ ID NO 1
LENGTH: 371
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Nearest "neighbor" = Streptomyces
OTHER INFORMATION: thermobaceous
US-09-321-981-1

Query Match 11.4%; Score 633.5; DB 4; Length 371;

```
Best Local Similarity 40.0%, Pred. No. 3.1e-37;
Matches 148; Conservative 52; Mismatches 131; Indels 39; Gaps 15;

QY 695 VAAVDAAGNTSAPSTPVDCETPEPNON-----GVTSVODGEYRVQTNENSSAQQCLTTINT 749
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 LASIALGALLTAAAPAO-----ANQOICDRYGTTIQD-RVYVQNNRMGTSAIQC--INV 65

QY 750 ATGAMTAVSTANSSGCTGAPATPSTIYKCHWNGCTTKNVGPIQISQIGSAVTSMTTQ 809
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 TGNGETIQAODGSVPTNAPKPSYPSYDCHGNCAPRTT-LPMKRISSIGSAPSSVSRY 124

QY 810 VSSGAYDAVADYDWTNSTPTTGTGQPNGETIMILWNSRGVOPGQSOTATGATVYAGHTWNV 869
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 TNGNYNAYADIDLPDTPRTNG-VNRTEIMTFNRYVGPQPGISPVGT-AHGGKRSMEW 182

QY 870 OGQOTSMKTIISVLTGATSIISNLDLKAIPADAARGSLINTSD-YLIDVEAGFEIWMQGO 928
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 TGSNSGNSNVISF-LAPSAISSWSFEDVKD-FYDQAVSHGILATPDWYLTISIOAGFEPEWEGT 240

QY 929 GLGSNSFSVSATSGG-----VACRATYYVNSDMGSGFATATYVINTGSRATNGWTY 979
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 GLAVNSFSSAVNAGGNGGCTPETPACQVSYSTHT-WPGEFTVDTTTNTGSTPVDMWEL 299

QY 980 AMSFGNGOTVTNMTALTQSGASVATNLTYSN-NVIOGQSTTFGNGSYSGTN-AAPT 1037
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 300 DFTLPAGHTVTSVWMLLSPASGAVTARSTGNSGRILANGTQSTGFCQTSAGAGFTAPA 359

QY 1038 -----LSCCT 1041
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 GARLNGTSCT 369

RESULT 7
US-09-216-295-21
; Sequence 21, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Collin
; TITLE OF INVENTION: No. 6268328el Variant EgIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216_295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Actinomycece 11Ag8
US-09-216-295-21

Query Match 10.2%, Score 564.5; DB 4; Length 312;
Best Local Similarity 41.3%, Pred. No. 1.9e-32;
Matches 129; Conservative 43; Mismatches 109; Indels 31; Gaps 12;

QY 695 VAAVDAAGNTSAPSTPVDCETPEPNON-----GVTSVODGEYRVQTNENSSAQQCLTTINT 749
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 LASIALGALLTAAAPAO-----ANQOICDRYGTTIQD-RVYVQNNRMGTSAIQC--INV 65

QY 750 ATGAMTAVSTANSSGCTGAPATPSTIYKCHWNGCTTKNVGPIQISQIGSAVTSMTTQ 809
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 TGNGETIQAODGSVPTNAPKPSYPSYDCHGNCAPRTT-LPMKRISSIGSAPSSVSRY 124

QY 810 VSSGAYDAVADYDWTNSTPTTGTGQPNGETIMILWNSRGVOPGQSOTATGATVYAGHTWNV 869
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 TNGNYNAYADIDLPDTPRTNG-VNRTEIMTFNRYVGPQPGISPVGT-AHGGKRSMEW 182

QY 870 OGQOTSMKTIISVLTGATSIISNLDLKAIPADAARGSLINTSD-YLIDVEAGFEIWMQGO 928
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 TGSNSGNSNVISF-LAPSAISSWSFEDVKD-FYDQAVSHGILATPDWYLTISIOAGFEPEWEGT 240

QY 929 GLGSNSFSVSATSGG-----VACRATYYVNSDMGSGFATATYVINTGSRATNGWTY 979
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 GLAVNSFSSAVNAGGNGGCTPETPACQVSYSTHT-WPGEFTVDTTTNTGSTPVDMWEL 299

QY 980 AMSFGNGOTVTNMTALTQSGASVATNLTYSN-NVIOGQSTTFGNGSYSGTN-AAPT 1037
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 300 DFTLPAGHTVTSVWMLLSPASGAVTARSTGNSGRILANGTQSTGFCQTSAGAGFTAPA 359

QY 1038 -----LSCCT 1041
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 GARLNGTSCT 369
```

```
Db 241 GLAVNSFSSAVNAGGNGGCTPETPACQVSYSTHT-WPGEFTVDTTTNTGSTPVDMWEL 299

QY 980 AMSFGNGOTVTN 991
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 300 DFTLPAGHTVTS 311

RESULT 8
US-09-118-319-9
; Sequence 9, Application US/09118319
; Patent No. 6114158
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Chen, Huizhong
; APPLICANT: Ljungdahl, Lars G.
; TITLE OF INVENTION: Oryzomyces Cellulase Celf Protein and Coding Sequences
; FILE REFERENCE: 33-98sequence listing
; CURRENT APPLICATION NUMBER: US/09/118_319
; CURRENT FILING DATE: 1998-07-17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 461
; TYPE: PRT
; ORGANISM: Fusarium oxysporum
US-09-118-319-9

Query Match 10.1%, Score 556.5; DB 3; Length 461;
Best Local Similarity 32.3%, Pred. No. 1.3e-31;
Matches 149; Conservative 71; Mismatches 167; Indels 75; Gaps 16;

QY 17 TTRMPAIRKRLAGVLAGAVSIAASIVPLAMQHPAIAATHVDNPYAGATFEVNPYMAQEV 76
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 70 TAAGPSSSTATKTATGSSSTTAGGSVTSA---PPAAS---DNFAYGVDIANNNTYRSEV 123

QY 77 OSEAA---NQTATLAAKRVVSTYSTAVWMDRIAINGVNGGPELTTYLDAALSQOQGT 133
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 NMLAVPKLSGAKATAAAVADVPSFO---WMDTYDHISLME---DTLADIRKANKAG- 174

QY 134 TREVEIETIYDLPGRDCAALANGELPATAAGLQTYENQYIDPIASISNPRYSLLRYT 193
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 GKTYAQFVYTDLPNRDCAALANGESLIDKDAANKYKA-YIAIKIGILQNN--YSOTKYL 231

QY 194 IIEPDSLPAVNTNMSIOACATAPPYEGIEYALTKLHAIPNVYIYMDAASHGWLGMVNN 253
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 VIEPDSLNLVNTNLVWDCAKAESAYKELTYVYAIKELN-LPVSVMYLDAGHGMWGMVNN 290

QY 254 ASGYQVEQKVLNASIGVNGIDGFTYNTANTYPLKEPMTATQVGGQPVESANFYQNNP 313
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 291 IGPAAKLYAQIYKDAKPSRVGLVTNVSNN-----GWLKSTKPDYTESNP 337

QY 314 DIDEADVAVDLSRLVYAAGFPSSIGMLIDTLNMGNGGNEPTSPSTANDVNTFVQSKID 373
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 338 NYDEQRT-INAFAPILAQDGMNVKFIYDQGRSG---KQPLG----- 375

QY 374 LRQHRGLMCONGAGIAGOPPOASPTDFPNALHDAVWIKPPESEDSGTSAAASPTTGKSKD 433
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 -QKAGDMCNAGTGFGLRPSNTNGD---ALADAEVWVKPGGESGCTS---DTSAAYTD 427

QY 434 PNCDDPTTYSYGLTNALPNSPIAGQWPPAOFDQLVANARRA 475
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 YHCG-----LDALKPAPEAGCTWQAYFEQLLDNANPS 459

RESULT 9
US-09-286-691-25
; Sequence 25, Application US/09286691
; Patent No. 6190189
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
```



```

; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96
; CURRENT APPLICATION NUMBER: US/09/286,691
; EARLIER FILING DATE: 1999-04-05
; CURRENT FILING DATE: 1996-10-04
; EARLIER APPLICATION NUMBER: US 60/027,883
; EARLIER FILING DATE: 1996-10-04
; EARLIER APPLICATION NUMBER: PCT US97/18008
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 25
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Fusarium oxysporum
; US-09-286-691-25

```

Query Match 10.0%; Score 555; DB 4; Length 360;

Best Local Similarity 33.7%; Pred. No. 1.1e-31; Mismatches 147; Indels 68; Gaps 14;

```

Matches 142; Conservative 64; Mismatches 147; Indels 68; Gaps 14;

OY 58 DNPYAGATFFVNPYMAOEVOSEAA---NOTNATLAAKMRVSTYSFAVMDRIAINGVN 114
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 3 DNPYAGVDLMANNYYRSEVMNLAVPKLSGAKATPAKAVADVSFQ---WMDTYDHISLME 59
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
OY 115 GGPGLTTLDAALSOOGCTPEVEIEIYIDLPGRDCAALASNGELPATPAAGLQTYETQYI 174
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 60 -----DTLADIRKANKAG-GKYAGQFVYDLPNRDCAAAASNGEYSLDKDANKYKA-YI 112
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
OY 175 DPASILSNPKYSLRIYITIEPDSLPAVNTNMSIQACATVAPYEGIEVALTKLHAIP 234
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 113 AKIKGILQN--YSPTKVILYIEPDSLNLVNLVNDKCAAEASAYKELTYAIKELN-LP 169
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
OY 235 NVYIYMDAAHSGMLGMPNNAAGYVOEVOKVLNASIGVNGIDGFTVNTANTYPLKEPFMTA 294
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 170 NYSMTLDHGGHGLGMPNNAIGPAKLYAQIKYKDGKPSRVKGLTVNVSNTN----- 220
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
OY 295 TOOVGOQPVESANFYQWNPDIADYAVDLYSRLVAAGFPSSIGMLDITLRNGMGPNP 354
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 221 ---GWLKSTRKPDYTESNPNTDEQRY-INAFAPLLAQEGSNVKTIVDQGRSG---KQP 271
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
OY 355 TGPSTADVTNTVYNOSKIDLRHRLGLMCNONGAGLGPPOASPTDFPNAHLDAVYWKIP 414
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 272 TG-----QKAQDGMCAKGTGFLRPSTWTGD---ALADAFVWVWPG 310
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
OY 415 GESDGTSAASDPPTGCKSDPMCDPTTYSYGVLTNALPNSPIAGQWFPQAQDOLVANARP 474
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 311 GESDGTSD---DTSAAARYDYHCG-----LDDALKPAPAGTWFQAYFRQLLDNANP 357
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
OY 475 A 475
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 358 S 358
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |

```

RESULT 10
US-09-687-147-25
Sequence 25, Application US/09687147
Patent No. 6268198

GENERAL INFORMATION:
APPLICANT: Li, Xin-Liang
APPLICANT: Ljungdahl, Lars G.
APPLICANT: Chen, Huizhong
TITLE OF INVENTION: Cellulases and Coding Sequences
FILE REFERENCE: 42-96A
CURRENT APPLICATION NUMBER: US/09/687,147
CURRENT FILING DATE: 2000-10-12
PRIOR APPLICATION NUMBER: US 60/027,883
PRIOR FILING DATE: 1996-10-04
PRIOR APPLICATION NUMBER: PCT US97/18008
PRIOR FILING DATE: 1997-10-03
PRIOR APPLICATION NUMBER: 09/286,691
PRIOR FILING DATE: 1999-04-05

```

; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 25
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Fusarium oxysporum
; US-09-687-147-25

```

Query Match 10.0%; Score 555; DB 4; Length 360;

Best Local Similarity 33.7%; Pred. No. 1.1e-31; Mismatches 147; Indels 68; Gaps 14;

Matches 142; Conservative 64; Mismatches 147; Indels 68; Gaps 14;

```

OY 58 DNPYAGATFFVNPYMAOEVOSEAA---NOTNATLAAKMRVSTYSFAVMDRIAINGVN 114
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 3 DNPYAGVDLMANNYYRSEVMNLAVPKLSGAKATPAKAVADVSFQ---WMDTYDHISLME 59
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
OY 115 GGPGLTTLDAALSOOGCTPEVEIEIYIDLPGRDCAALASNGELPATPAAGLQTYETQYI 174
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 60 -----DTLADIRKANKAG-GKYAGQFVYDLPNRDCAAAASNGEYSLDKDANKYKA-YI 112
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
OY 175 DPASILSNPKYSLRIYITIEPDSLPAVNTNMSIQACATVAPYEGIEVALTKLHAIP 234
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 113 AKIKGILQN--YSPTKVILYIEPDSLNLVNLVNDKCAAEASAYKELTYAIKELN-LP 169
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
OY 235 NVYIYMDAAHSGMLGMPNNAAGYVOEVOKVLNASIGVNGIDGFTVNTANTYPLKEPFMTA 294
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 170 NYSMTLDHGGHGLGMPNNAIGPAKLYAQIKYKDGKPSRVKGLTVNVSNTN----- 220
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
OY 295 TOOVGOQPVESANFYQWNPDIADYAVDLYSRLVAAGFPSSIGMLDITLRNGMGPNP 354
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 221 ---GWLKSTRKPDYTESNPNTDEQRY-INAFAPLLAQEGSNVKTIVDQGRSG---KQP 271
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
OY 355 TGPSTADVTNTVYNOSKIDLRHRLGLMCNONGAGLGPPOASPTDFPNAHLDAVYWKIP 414
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 272 TG-----QKAQDGMCAKGTGFLRPSTWTGD---ALADAFVWVWPG 310
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
OY 415 GESDGTSAASDPPTGCKSDPMCDPTTYSYGVLTNALPNSPIAGQWFPQAQDOLVANARP 474
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 311 GESDGTSD---DTSAAARYDYHCG-----LDDALKPAPAGTWFQAYFRQLLDNANP 357
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
OY 475 A 475
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |
DB 358 S 358
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |

```

RESULT 11
US-09-118-319-8
Sequence 8, Application US/09118319
Patent No. 6114158

GENERAL INFORMATION:
APPLICANT: Li, Xin-Liang
APPLICANT: Ljungdahl, Lars G.
APPLICANT: Chen, Huizhong
TITLE OF INVENTION: Orp1omycetes Cellulase CelF protein and Coding Sequences
FILE REFERENCE: 33-98sequence listing
CURRENT APPLICATION NUMBER: US/09/118,319
CURRENT FILING DATE: 1998-07-17
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 8
LENGTH: 470
TYPE: PRT
ORGANISM: Trichoderma reesei
US-09-118-319-8

Query Match 9.9%; Score 548.5; DB 3; Length 470;
Best Local Similarity 31.6%; Pred. No. 4.8e-31; Indels 109; Gaps 17;
Matches 158; Conservative 56; Mismatches 177; Indels 109; Gaps 17;

```

OY 7 GGRNCRKQR-----GTRRM-PAISKRRLRAGVLAGAVSIASIVPL 45
| | | | | : | : | : | : | | | | : | : | | | : | : | | | : | : |

```

```
Db 47 SSGTCVSNDDYYSQCLPGAASSSSSTRASTTRVSPPTSSSSATPPPG--STTTAVPV 104
QY 46 AMOHPAIAATHVDNPYAGATFFVNPYMAQEOVSEANQTNATLAAMKRVSTYSTAVMD 105
Db 105 G-----SGTATYSGNPEVGYTPMANAYASEVSLAIPSLTGAMATAAAVAKVSPFMD 160
QY 106 -----RAATNGVNGSGPLTYYLDALSSQOQGTPEVIEIYIYDLPGRDCAALAS 155
Db 161 TUDKTPPLMQTLADITKANGG--NYAG-----QFVYDLPDRDCAALAS 204
QY 156 NGELPATAAGTQYETQYIDIASILSNPKYSRLIVTIIEDSLPNAVTNMSIQACATA 215
Db 205 NGEYSIADGVAKK--NYIDITRQIV--VEYSDIRTLVIEPDSLANTYTNLGTPRCANA 261
QY 216 VPEYEGICETALTKLHAIDNPVYIYMDAHSGLWGPNNASGYVOEQVYKVLASIGVNGID 275
Db 262 OSATYECIYIAVYQOLN-LPNVAMYIDAGHAGWIGPANODPAQOLFANYYKNASSPRALR 320
QY 276 GFVTNTANTPLKEPMTATQOVGPVESANFYQNNPIDEDADYAVDLXSLVAAFGPS 335
Db 321 GLATNVANTN-----GWNITSPSTQGNNAVYNEKLYTHAIGPLLANHGW-S 366
QY 336 SIGMLIDTLRNGMGNEPTGPSTATDVNTFVNSQKIDLRHGLMCNONGAGLQGPQA 395
Db 367 NAFETIDGRRG---KQPTG-----QQQKGMKCNVIGTGFGRPSA 404
QY 396 SPDPFPNAHLDAVYWKIPGESDGTSAASDPPTGKKSDPMCDPTYYTSYGVLTNALPNSP 455
Db 405 NMGD---SLIDSFVWVWPKGECDDGTSSSAP---RFDSHC-----ALPDALQAP 448
QY 456 IAGOWEPAQFDOLVANARPA 475
Db 449 QAGAMFOATFVOLLTNANPS 468
```

```
RESULT 12
US-09-142-759-1
; Sequence 1, Application US/09142759A
; Patent No. 6127160
; GENERAL INFORMATION:
; APPLICANT: YAMANOBE, Takashi
; APPLICANT: WATANABE, Manabu
; APPLICANT: HAMAYA, Toru
; APPLICANT: SUMIDA, Naomi
; APPLICANT: AOYAGI, Kaoru
; APPLICANT: MORAKAMI, Takeshi
; TITLE OF INVENTION: PROTEIN HAVING CELLULASE ACTIVITY AND PROCESS FOR
; FILE REFERENCE: 051673
; CURRENT APPLICATION NUMBER: US/09/142,759A
; EARLIER FILING DATE: 1998-09-14
; EARLIER APPLICATION NUMBER: PCT/JP97/00824
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: P. HEI-8-84479
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Acromonium cellulosilyticus
US-09-142-759-1
```

```
Query Match 9.8%; Score 544; DB 3; Length 457;
Best Local Similarity 32.5%; Pred. No. 9.5e-31;
Matches 151; Conservative 60; Mismatches 171; Indels 82; Gaps 16;
QY 16 GTTRPAISAKRLRACVAGVASTIASIVPLAQHPAIAATHVDNPYAGATFFVNPYMAQ 75
Db 74 GTT-SPTTTTAKSTATTATTAASG-----NPSGYQLVANPYSSB 115
QY 76 VQSEANQTNATLAAMKRVSTYSTAVMDRIAINGVNGSGPLTYYL-DALSSQOQGT 134
```

```
Db 116 VHTLAIPSLTGLAAATRAAETIPSFVWLDTAKY-----PTMGTYLIANEAANKAGAS 169
QY 135 PEVIEI-VIYDLPGRDCAALASNGELPATAGTQYETQYIDIASILSNPKYSRLIYV 193
Db 170 PPIAGIFVYDLPDRCAALASNGELTYVANNGVANKA-YIDISYAQLR--AYPDVHTIL 226
QY 194 IIEPDSLPAVNTNMSIQACATAVPEYEGIEYALTKLHAIPNVYIYMDAHSGLWGPNN 253
Db 227 IIEPDSLAMNVTNLSTAKCAEAQASAYEECVNVALNLN-LANVAMYIDAGHAGWGSAN 285
QY 254 ASGYVOEQVYKVLASIGVNGIDGFYNTANTYTPKPEPMTAQOVGGQGVSEANFYQNNP 313
Db 286 LSPAAQLFATVYKNASAPASLGLATNVANYN-----AMSSISPP-----SYTSGDS 332
QY 314 DIDEDAYAVDLXSLVAAAFPSISGLIDTLRNGMGNEPTGPSTATDVNTFVNSQKID 373
Db 333 NYDEKILYNALSPILTSKNMPNA-HFIMDTSRNG---VQPT----- 369
QY 374 LQHRGLMCNONGAGLQGPQASPTDFPNAHLDAVYWKIPGESDGTSAASDPPTGKKSD 433
Db 370 KOQAMGDMCNVIGTGFGRVQPTNTGTD---PLEDAFVWVWPKGESDGTSSNS---ATRYD 422
QY 434 PMCDPTYYTSYGVLTNALPNSPVIAQWFPAPQFDOLVANARPAVY 477
Db 423 FRCG-----YSDALQAPAEAGTWQATFVOLLTNANPALV 457
```

```
RESULT 13
US-09-286-691-24
; Sequence 24, Application US/09286691
; Patent No. 6190189
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96
; CURRENT APPLICATION NUMBER: US/09/286,691
; EARLIER FILING DATE: 1999-04-05
; EARLIER APPLICATION NUMBER: US 60/027,883
; EARLIER FILING DATE: 1996-10-04
; EARLIER APPLICATION NUMBER: PCT US97/18008
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-286-691-24
```

```
Query Match 9.7%; Score 538; DB 4; Length 360;
Best Local Similarity 33.5%; Pred. No. 1.8e-30;
Matches 143; Conservative 50; Mismatches 152; Indels 82; Gaps 13;
QY 59 NPYAGATFFVNPYMAQEOVSEANQTNATLAAMKRVSTYSTAVVMD-----RIA 108
Db 4 NPEVGYTPMANAYVYASEVSLAIPSLTGAMATAAAVAVSPFMDITDITKTPMLMQTLA 63
QY 109 AINGVNGSGPLTYYLDALSSQOQGTPEVIEIYIYDLPGRDCAALASNGELPATAGIQT 168
Db 64 DIRTKANGG--NYAG-----QFVYDLPDRDCAALASNGEYSIADGVAK 107
QY 169 YETQYIDPLASILSNPKYSRLIVTIIEDSLPNAVTNMSIQACATAVYIYEDGIEYALT 228
Db 108 YK-NYIDITRQIV--VEYSDIRTLVIEPDSLANTYTNLGTPRCANAQASAYIECTINAVT 164
QY 229 KLHAIDPNVYIYMDAHSGLWGPNNASGYVOEQVYKVLASIGVNGIDGFYNTANTYTPK 288
Db 165 QLN-LPNVAMYIDAGHAGWIGPANODPAQOLFANYYKNASSPRALRLGLATNVANYN--- 220
```

```
QY 289 EPFMTATQOYGGPVESEANFYQWNPDIIDEADYAVDLYSRLYVAGFPSSIGMLIDTLRNGW 348
DB 221 -----GWNITSPSPSTQGNAYVNEKLYIHAIIGPLANHGWSNAFFITDQGRSG- 268
QY 349 GGPNEPTGPSTATDVNTFVNOSKIDLRHRIAGMCNONGAGLGOPPOASPTDFPNAHLDAY 408
DB 269 ---KQPTG-----OOQMGDKMNVICTGFGIRPSANTGD---SLDSF 304
QY 409 VWKPGESDGTSAASDPTTGKKSPPMCDPTTYSYGLVTALNPNSPIATGQMPFAQDFDL 468
DB 305 VWKPGEGCDGTSDSSAP-----RFDSHC-----ALPDALQPAQAGAMFQAYFYVL 351
QY 469 VANARPA 475
DB 352 LTNANPS 358

RESULT 14
US-09-687-147-24
; Sequence 24, Application US/09687147
; Patent No. 6268198
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Chen, Huizhong
; APPLICANT: Ljungdahl, Lars G.
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96a
; CURRENT APPLICATION NUMBER: US/09/687,147
; PRIOR FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/027,883
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: PCT US97/18008
; PRIOR FILING DATE: 1997-10-03
; PRIOR APPLICATION NUMBER: 09/286,691
; PRIOR FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 24
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-687-147-24

Query Match 9.7%; Score 538; DB 4; Length 360.
Best Local Similarity 33.5%; Pred. No. 1.8e-30;
Matches 143; Conservative 50; Mismatches 152; Indels 82; Gaps 13;

QY 59 NPVAGATFEVNPVMAQVQSEANQTNATLAAKMRVSTYSTAVWMD-----RIA 108
DB 4 NPEVGVTPMANAYYASEVSSLAIPSLTGAMATAAAVAKVPSFWMIDTLDTKPLMEQTILA 63
QY 109 AINGVNGSGPLTTLDAALSOOGCTTPEVIEIYIDLPGRCALASNGELPATPAAGLOT 168
DB 64 DIRPANNNGG--NYAG-----QFVYIDLPRDCAALASNGEYSIADGGVAK 107
QY 169 YETQYIDPLASILSNPKYSLRIYITIEPDSLPAVNTNMSIQACATAPVPEEGIEYALT 228
DB 108 YK-NYIDTIRQIV--VEYSDIRTLVIEPDSLANTLNTLGPCKANQASATLECIANTAVT 164
QY 229 KLAHAIPIVYIYMDAASHGWLGMPPNNAAGVQVQKVLNASTGVNGIDGFTVNTANTPLK 288
DB 165 QLN-LPYNVAMYLDAHGHWLGMPPANODPAAQLFANVYKNASSPRALRGLATNVANYN--- 220
QY 289 EPFMTATQOYGGPVESEANFYQWNPDIIDEADYAVDLYSRLYVAGFPSSIGMLIDTLRNGW 348
DB 221 -----GWNITSPSPSTQGNAYVNEKLYIHAIIGPLANHGWSNAFFITDQGRSG- 268
QY 349 GGPNEPTGPSTATDVNTFVNOSKIDLRHRIAGMCNONGAGLGOPPOASPTDFPNAHLDAY 408
DB 269 ---KQPTG-----OOQMGDKMNVICTGFGIRPSANTGD---SLDSF 304
QY 409 VWKPGESDGTSAASDPTTGKKSPPMCDPTTYSYGLVTALNPNSPIATGQMPFAQDFDL 468
```

```
DB 305 VWKPGEGCDGTSDSSAP-----RFDSHC-----ALPDALQPAQAGAMFQAYFYVL 351
QY 469 VANARPA 475
DB 352 LTNANPS 358

RESULT 15
US-08-169-948B-12
; Sequence 12, Application US/08169948B
; Patent No. 5861271
; GENERAL INFORMATION:
; APPLICANT: Fowler, Timothy
; APPLICANT: Ward, Michael
; APPLICANT: Clarkson, Kathleen
; APPLICANT: Collier, Katherine
; APPLICANT: Larens, Edmund
; TITLE OF INVENTION: No. 5861271el Cellulase Enzymes and Systems
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk-
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,948B
; FILING DATE: DEC 17 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A.
; REGISTRATION NUMBER: 33,401
; REFERENCE/DOCKET NUMBER: GC226
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7536
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 365 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-169-948B-12

Query Match 9.7%; Score 538; DB 2; Length 365;
Best Local Similarity 33.3%; Pred. No. 1.8e-30;
Matches 144; Conservative 51; Mismatches 155; Indels 82; Gaps 13;

QY 54 ATHVQNPVAGATFEVNPVMAQVQSEANQTNATLAAKMRVSTYSTAVWMD----- 105
DB 4 ATYSGNPFVGTVMANAYYASEVSSLAIPSLTGAMATAAAVAKVPSFWMIDTLDTKPLM 63
QY 106 --RIAINGVNGSGPLTTLDAALSOOGCTTPEVIEIYIDLPGRCALASNGELPATPA 163
DB 64 EGTLDITRANKNG--NYAG-----QFVYIDLPRDCAALASNGEYSIAD 107
QY 164 AGLOTYETQYIDPLASILSNPKYSLRIYITIEPDSLPAVNTNMSIQACATAPVPEEGCI 223
DB 108 GGVARYK-NYIDTIRQIV--VEYSDIRTLVIEPDSLANTLNTLGPCKANQASATLECI 164
QY 224 EYALTKLAHAIPIVYIYMDAASHGWLGMPPNNAAGVQVQKVLNASTGVNGIDGFTVNTAN 283
DB 165 NYAVTQLN-LPYNVAMYLDAHGHWLGMPPANODPAAQLFANVYKNASSPRALRGLATNVAN 223
```

```

Oy 284 YTPLEPFMTATQYGGOPVESANFYQWNPDIADYAVDLXSRLVAGFSSIGMLIDT 343
Db 224 YN-----GWNITSPSTYQGNVAVYNEKLYIHAIGPLANHGM-SNAFFITDQ 269
Oy 344 LKNGWGPNPTGPSTATDVNTFVNQSKIDLRQHRGLMCNONGAGIGPPQASPTDFPNA 403
Db 270 GRS-----KQPTG-----QQMGDMCNVIGTGFGIRPSANTGD---S 304
Oy 404 HLDAYVWIKPPGESDGTSAASDPTTGKKSDPWCDDPTTSTYGVLTNALPNSPIAGOWPFA 463
Db 305 LLDSEVWVKPGGECDDGTSDSSAP---RDSHC-----ALPDALQAPAPQAGAMFOA 351
Oy 464 QFDOLVANARPA 475
Db 352 YFVQULTNANPS 363

```

Search completed: August 29, 2002, 16:22:37
 Job time: 367 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:21:46 ; Search time 127 Seconds
(without alignments)
202.032 Million cell updates/sec

Title: US-09-917-384-7

Perfect score: 1244

Sequence: 1 DCTPCPNQNGVTYVDGEYR.....IMWGGGLGNSNFSVSTSG 231

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_032802.*
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT.*
23: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	472.5	38.0	371	20	AAV06367
2	472.5	38.0	371	20	AAV08473
3	472.5	38.0	371	21	AAV08473
4	472.5	38.0	371	21	AAV08473
5	472.5	38.0	371	21	AAV08473
6	472.5	38.0	371	21	AAV08473
7	472.5	38.0	371	21	AAV08473
8	472.5	38.0	371	21	AAV08473
9	472.5	38.0	371	21	AAV08473
10	472.5	38.0	371	21	AAV08473
11	472.5	38.0	371	21	AAV08473

12	407.5	32.8	260	21	AAV84347
13	260.5	20.9	259	20	AAV06351
14	260.5	20.9	259	21	AAV06351
15	260.5	20.9	259	21	AAV06351
16	260.5	20.9	259	21	AAV06351
17	260.5	20.9	259	21	AAV06351
18	260.5	20.9	259	21	AAV06351
19	260.5	20.9	259	21	AAV06351
20	260.5	20.9	259	21	AAV06351
21	260.5	20.9	259	21	AAV06351
22	260.5	20.9	259	21	AAV06351
23	260.5	20.9	259	21	AAV06351
24	260.5	20.9	259	21	AAV06351
25	260.5	20.9	259	21	AAV06351
26	260.5	20.9	259	21	AAV06351
27	260.5	20.9	259	21	AAV06351
28	260.5	20.9	259	21	AAV06351
29	260.5	20.9	259	21	AAV06351
30	260.5	20.9	259	21	AAV06351
31	260.5	20.9	259	21	AAV06351
32	260.5	20.9	259	21	AAV06351
33	260.5	20.9	259	21	AAV06351
34	260.5	20.9	259	21	AAV06351
35	260.5	20.9	259	21	AAV06351
36	260.5	20.9	259	21	AAV06351
37	260.5	20.9	259	21	AAV06351
38	260.5	20.9	259	21	AAV06351
39	260.5	20.9	259	21	AAV06351
40	260.5	20.9	259	21	AAV06351
41	260.5	20.9	259	21	AAV06351
42	260.5	20.9	259	21	AAV06351
43	260.5	20.9	259	21	AAV06351
44	260.5	20.9	259	21	AAV06351
45	260.5	20.9	259	21	AAV06351

ALIGNMENTS

RESULT	1
AAV06367	standard; Protein; 371 AA.
ID	AAV06367
XX	AAV06367
AC	AAV06367
XX	06-SEP-1999 (first entry)
DE	Streptomyces sp. EGIIT-like cellulase.
XX	Cellulase; endoglucanase; EGIIT; textile; feed additive; baking;
KW	food processing; grain wet milling; pulp; paper.
XX	Streptomyces sp.
OS	Streptomyces sp.
XX	W09931255-A2.
PN	24-JUN-1999.
XX	14-DEC-1998; 98WO-US26552.
XX	16-DEC-1997; 97US-0991720.
PR	(GENEV) GENENCOR INT INC.
PA	Bower BS, Fowler T, Phillips JT;
XX	WPI; 1999-395187/33.
DR	EGIT like cellulase
XX	Example; Fig 6; 47pp; English.
PS	The present polypeptide represents a full-length sequence of a
XX	
CC	

Amino acid sequenc
Aspergillus aculea
Aspergillus aculea
Amino acid sequenc
Streptomyces sp. E
Glicolactium roseum
Glicolactium roseum
Amino acid sequenc
M. cinnamomea xyl
Bacillus lichenif
Streptomyces livid
Erythrina carotovora
Amino acid sequenc
Emeticella deserta
Actinomyces I1AC8
Amino acid sequenc
Fusarium equiseti
Fusarium equiseti
Aspergillus kawach
Aspergillus kawach
Amino acid sequenc
Aspergillus niger
Cellulase polypt
Chaetomium brasill
Amino acid sequenc
Tiarosporella phas
T. phaseolina xyl
Glicolactium roseum
Glicolactium roseum
Amino acid sequenc
Glicolactium roseum

PS Disclosure: Page 64-65; 72pp; English.

Query Match 38.08; Score 472.5; DB 21; Length 371;

```

QY      10  GYTSVQDGEYRYQVINEWMSAAOCLITFATGANTVSTIANSSGCTGGAPNPYISYKCH 69
       11  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      40  gttlqgd-lyvvannrgwtsatqc--invtngteltlaagdvpcunqapksypsvyqch 96
QY      70  WGCCTKTKNGMEIOISQISAVTSMSTQVSSGAYDAVDIMTNSPTTGGPNGTEIMI 129
       11  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      97  ygncaprtclpmrtlsisgsapsvsylytqgyvnaeydwlqptptng-vrrtclmi 154
QY      130  WLNSSRGVOPFGSQIRATGTVTAQHMMWNWQGOQTSWKIISVILPFGASINLDAIFA 189
       11  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      155  wfmrnpvvpqplspayql-ahvgaisvewtvgngsndrlsf-lapsalsavsfaykd-fv 211

```

RESULT	5
AA67496	
ID	AA67496 standard; protein; 371 AA
XX	

AC AAY67496;
XX
XX
DT 19-MAY-2000 (first entry)
XX
XX Actinomycetes cellulase protein sequence.
DE Actinomycetes: detergent; feed additive; textile treatment.
XX Cellulase; Actinomycetes: detergent; feed additive; textile treatment.
KW pulp; paper.
RV

OS	Streptomyces sp.	Location/Qualifiers
XX	key	1..27
FH	peptide	/note= "signal sequence"
FI		28..372
FI	Protein	/note= "mature protein"
FT		
XX		
PN	W0200009707-AL.	
XX		
PD	24-FEB-2000.	
XX		
PF	28-MAY-1999;	99WO-US1197L.
XX		
PR	24-JUN-1998;	98US-O104308. 204640.

PR 18-NOV-1998; 98W0-0524845
PR 28-MAY-1999; 9905-0321981
XX
XX (GEMV) GENENCOR INT INC.
XX

XX Novel variant endoglucanase III-like cellulases with improved
PT surfactant stability and resistance to temperature stress, useful for
PT textile processing or cleaning, treating wood pulp, food and grain, and
PS reducing biomass to glucose

Disclousure; Page 65-66; 73pp; English.

CC The present sequence represents an endoglucanase III (EgIII)-like
CC cellulase. The cellulase has homology to the Trichoderma reesei EgIII
CC protein. The variant cellulases have improved temperature stability,
CC and improved surfactant stability. The variant cellulases and
CC compositions containing them are used in textile processing or cleaning,
CC e.g. stonewashing of indigo dyed denim, and modifying the texture, feel
CC or appearance of cellulose containing fabrics (e.g. improving fabric
CC smoothness or removing pills and fibrils). The compositions may also be
CC used for the removal of immature or dead cotton from cellulosic fibres
CC or fabric, which can cause uneven dyeing. The cellulase may also be used
CC in a detergent composition for washing laundry and dishes and in the
CC treatment of animal feed, wood pulp, paper, non-animal foods and grains.
CC The enzymes may also be used in the reduction of biomass to glucose.

XX Sequence 429 AA:

Query Match 32.9%; Score 409; DB 21; Length 429;
Best Local Similarity 41.6%; Pred. No. 1.8e-27;
Matches 91; Conservative 32; Mismatches 90; Indels 6; Gaps 6;

OY 13 SVQDEYRVQNTENMNSSAOCLTINTAGMTVSTANFSGGTGAPATYPSIKGCHWG 72
DB 99 svfngghyqnmrwgstapcvta-tdty-frvtgdsgapnpgakspysvfnghychn 156
OY 73 CTTKNVGMPIQISQISAVTSMSTTOVSSGADVADVTNSTPTTGQPNGTETIMWLN 132
DB 157 c-spqtdlpvridtvaapsisysygfvdgavyvnyasydiwldpccardg-vnqetelmwfn 214
OY 133 SRGGOVPFSGQATGVTVAGHTVMWMOGQOTSMKIISYVLTPEATGISNLDKAITRADAA 192
DB 215 rvgpdpdpispygt-asvvggrtewvsgngndvlsiv-apsaisqwsldvmdfvravt 272
OY 193 ARGSLNTSDYLDVEAGFEIMOGGQIGSNSFSVSTSG 231
DB 273 arglaendwyltsvqagfepwqngaglavnsfstvctg 311

RESULT 10

AA06369
ID AAY06369 standard; Protein; 260 AA.

XX AAY06369;

XX 06-SEP-1999 (first entry)

XX Rhodothermus marinus EgIII-like cellulase.

XX Cellulase: endoglucanase; EgIII; textile; feed additive; baking;

XX food processing; grain wet milling; pulp; paper.

XX Rhodothermus marinus.

XX WO931255-A2.

XX 24-JUN-1999.

XX 14-DEC-1998; 98WO-US26552.

XX 16-DEC-1997; 97US-0991720.

XX (GENV) GENENCOR INT INC.

XX Bower BS, Fowler T, Phillips JT;

DR WPI; 1999-395187/33.

XX EgIII like cellulase

XX Example; Fig 6; 47pp; English.

XX The present polypeptide represents a full-length sequence of a
CC novel EgIII-like cellulase of Rhodothermus marinus. It was
CC deduced from a gene sequence isolated from genomic DNA using PCR
CC primers (see AAX59180-91) based on conserved motifs (see AAY06325-29)
CC of Trichoderma reesei EgIII cellulase and related enzymes. PCR
CC has been used to identify novel EgIII-like enzymes, including the
CC present protein, from bacterial and fungal sources (see AAY06331-70).
CC The sequence shows homology to T. reesei EgIII (see AAY06330). Also
CC provided by the invention are vectors, host cells and methods
CC for the recombinant production of such enzymes, which can be used
CC in the treatment of cellulose-containing textiles, as feed
CC additives, in the treatment of wood pulp, in the reduction of
CC biomass to glucose, in the treatment of indigo dyed denim, or
CC as laundry detergent components (all claimed).

XX Sequence 260 AA:

Query Match 32.8%; Score 407.5; DB 20; Length 260;
Best Local Similarity 40.1%; Pred. No. 1.3e-27;
Matches 89; Conservative 30; Mismatches 84; Indels 19; Gaps 7;

OY 14 VQDEYRVQNTENMNSSAOCLTINTAGMTVSTANFSGGTGAPATYPSIKGCHWG-- 71
DB 50 vaggrrvrvlnwvgaetaqlvleqtnftlcradhdgmn--vaaypalyfychwapa 107
OY 72 ----NCTKNVGMPIQISQISAVTSMSTTOVSSGADVADVTNSTPTTGQPNGTET 126
DB 108 rairdcaar-agavvrrahel-----dvrlptlgrwnaaydiwfsfpvtnsgnygsgae 159
OY 127 IMTWLNRGGOVPFSGQATGVTVAGHTVMWMOGQOTSMKIISYVLTPEATGISNLDKAI 186
DB 160 lmlwlnmgvgmpgsrval-velagatwewvya-dwdwnylayrrtptltsvseidlka 217
OY 187 IFADAAARGSLNTSDYLDVEAGFEIMOGGQIGSNSFSVS 228
DB 218 flddavaraylirpewylhavevtfelweggaglrtdafstv 259

RESULT 11

AAB1482
ID AAB1482 standard; Protein; 260 AA.

XX AAB1482;

XX 21-NOV-2000 (first entry)

XX Emeritcella desertoru EgIII-like cellulase.

XX Emeritcella desertoru; Trichoderma reesei; endoglucanase III; EgIII;

XX cellulase; mutant; enzyme stability; textile treatment;

XX wood pulp treatment; feed additive; detergent.

XX Emeritcella desertoru.

XX WO200037614-A2.

XX 29-JUN-2000.

XX 12-NOV-1999; 99WO-US26704.

XX 18-DEC-1998; 98US-0216295.

XX (GENV) GENENCOR INT INC.

XX Mitchinson C, Wendt DJ;

[illegible]

```

XX      (GENV ) GENENCOR INT INC.
PA
XX
XX      Mitchinson C, Wendt DJ;
PI
XX
XX      WPL; 2000-482483/42.
DR
XX
XX
XX      Novel endoglucanase III or endoglucanase III-like cellulase useful for
PT      treating textiles and wood pulp comprises a substitution or deletion at
PT      specified positions in the wild form of endoglucanase III
XX
XX
PS      Example 1; Fig 3; 52pp; English.
XX
CC      The present sequence is a cellulase related to endoglucanase III (EgIII)
CC      from Trichoderma reesei. EgIII-like genes were isolated from genomic DNA
CC      libraries constructed from various microorganisms by PCR. The isolated
CC      clones showed significant homology to EgIII from T. reesei. Certain
CC      substitution and deletion mutations have been incorporated into EgIII and
CC      EgIII-like cellulases to produce variant enzymes with improved stability,
CC      e.g. increased resistance to temperature stress. The mutants may be used
CC      in textile and wood pulp treatment, as a feed additive, and for reducing
CC      biomass to glucose. They are also useful for stonewashing or indigo dyed
CC      denim and as an agent in laundry and dish detergents.
XX
SQ      Sequence      259 AA;
XX
XX
XX      Query Match      20.9%; Score 260.5; DB 21; Length 259;
XX      Best Local Similarity 33.9%; Pred. No. 8.4e-15;
XX      Matches 82; Conservative 21; Mismatches 96; Indels 43; Gaps 10
XX
QY      17 GERVQVQTEBN---SSAQCCLTINTATGAMV-STA-NEGSGTGAPATYPSIYKCHWG 71
DB      31 gvylinmlwgkdgagsqgctlvnsasagstwtstkwnsgensvks-----yan 81
XX
QY      72 NCTKNKNGMFIQISQIGSAVTSMTQVSSGAYDAVDLWNTSTRTTTCGQENGEMIMWL 131
DB      82 sglfnnkklvsgisqj-ptlarsydnctgira-dvaaydlftaadlnhwtsgdyelmwl 139
XX
QY      132 NSRNGVQPEFSQATGVTYAGHTWANNWQOQTSMKRTISVLTLPGATGISNLDLKAIFDA 191
DB      140 airtgvgvbpisqgat-ctvdgqgtwelwtyangansqktystfvaprltstfg-dvndfky 197
XX
QY      192 AA-RGSLINTSDYLIV------EAGFEIMOGGOGGLSNSPSV 226
DB      198 tqnqgfpassqylltqfgrptgpgpalsvsnwsasvqgagfepwngaglavnsfss 257
XX
QY      227 SV 228
DB      258 tv 259
XX
XX
XX      RESULT 15
XX      ID      AAY84329 standard; Protein; 259 AA.
XX      AAY84329;
XX
XX      12-JUL-2000 (first entry)
XX
XX      Amino acid sequence of an endoglucanase III (EgIII)-like cellulase.
XX
XX      Endoglucanase III; EgIII; EgIII-like cellulase; surfactant stability;
XX      cellulase; textile processing; textile cleaning; stonewashing;
XX      indigo dyed denim; cellulose containing fabric; fabric smoothness;
XX      pill removal; fibril removal; cotton; cellulosic fibre; dyeing; detergent;
XX      animal feed; wood pulp; paper; grain; biomass reduction; glucose.
XX
XX      Aspergillus aculeatus.
XX
XX      WO2000014208-A1.
XX
XX      16-MAR-2000.
XX
XX

```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:24:14 ; Search time 72.78 seconds
(without alignments)
304.983 Million cell updates/sec

Title: US-09-917-384-7
Perfect score: 1244
Sequence: 1 DCTPGPNQGVTSVQDEYR.....IMOGGGLGSNSFSVTS 231

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: p1r1.*
2: p1r2.*
3: p1r3.*
4: p1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	436.5	35.1	382	JC2571	cellulase (EC 3.2.1.4)
2	235.5	18.9	151	H70895	hypothetical prote
3	234	18.8	237	S12610	cellulase (EC 3.2.1.4)
4	215	17.3	264	JU0328	cellulase (EC 3.2.1.4)
5	184.5	14.8	239	S55931	cellulase (EC 3.2.1.4)
6	133	10.7	334	G90360	endoglucanase prec
7	139	10.4	1286	S28634	adhesin AIDA-I pre
8	138	10.3	332	G90291	endoglucanase prec
9	112	9.0	1025	G81722	polymorphic membra
10	108	8.7	1275	T33369	hypothetical prote
11	108	8.7	1777	T34360	hypothetical prote
12	105.5	8.5	258	H72240	endoglucanase - Th
13	105.5	8.5	958	A82583	conserved hypotet
14	105	8.4	1036	A29832	HPI layer surface
15	104	8.4	747	B47093	cellulase (EC 3.2.1.4)
16	100	8.0	4776	E95206	cell wall surface
17	98.5	7.9	526	A56573	nuclear pore compl
18	98	7.9	1390	T18883	hypothetical prote
19	98	7.9	1749	S75138	hypothetical prote
20	98	7.9	3075	S14458	laminin alpha-1 ch
21	97.5	7.8	562	AE1335	probable peptidog
22	97.5	7.8	5188	B85547	probable RTX fami
23	97.5	7.8	5291	E90696	hypothetical prote
24	97.5	7.8	13288	T03099	mucin, submaxillar
25	97	7.8	985	D89852	fibrogen-binding
26	96.5	7.8	948	C75265	hexagonally packe
27	96.5	7.8	1004	T00046	surface layer prot
28	96.5	7.8	1076	A35622	nuclear pore prote
29	96.5	7.8	1649	C86822	hypothetical prote

30	96.5	7.8	3716	2	E70969	probable ppe prote
31	96	7.7	479	2	C90457	cytochrome b558/56
32	96	7.7	1441	2	A86685	phage pil prote
33	96	7.7	1475	2	S42718	nuclear pore compl
34	95.5	7.7	839	2	H97758	outer membrane ass
35	95	7.6	211	2	C37491	hypothetical prote
36	95	7.6	221	2	UC7307	endo-1,4-beta-xylo
37	95	7.6	666	2	A42296	lysosyme 2 (EC 3.2
38	95	7.6	1345	2	H90975	hypothetical prote
39	95	7.6	1461	2	E90696	hypothetical prote
40	95	7.6	1461	2	A85547	hypothetical prote
41	95	7.6	2232	2	T34434	probable invasin 2
42	95	7.6	2660	2	E85822	hypothetical prote
43	94.5	7.6	1251	2	T21389	immunodominant ant
44	94	7.6	233	2	B90062	hypothetical prote
45	94	7.6	657	2	S25184	cspl protein - Cor

ALIGNMENTS

RESULT 1
JC2571
cellulase (EC 3.2.1.4) precursor - Streptomyces rochei (strain A2)
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase
C:Species: Streptomyces rochei
C:Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 22-Oct-1999
C:Accession: JC2571; S34392
R:Porto, B.; Hanhart, E.; Irdani, T.; Iqbal, M.; McCarthy, A.J.; Mastromei, G.
Gene 148, 119-124, 1994
A:Title: Characterization and sequence analysis of a Streptomyces rochei A2 endoglucanase
A:Reference number: JC2571; MUID:95011642
A:Accession: JC2571
A:Molecule type: DNA
A:Residues: 1-382 <PER>
A:Cross-references: EMBL:X73953; NID:g393391; PIDN:CA52139.1; PID:g393392
A>Note: this cellulolytic strain was isolated from the gut of termites
C:Genetics:
A:Gene: egis
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Superfamily: bacterial cellulose-binding domain homology
C:Keywords: glycosidase; hydrolyase; polysaccharide degradation
F:1-37/Domain: signal sequence #status predicted <SID>
F:38-382/Product: endoglucanase #status predicted <MAT>
F:219-380/Domain: bacterial cellulose-binding domain homology <BCB>
F:280-379/Disulfide bonds: #status predicted

Query Match 35.1%; Score 436.5; DB 2; Length 382;
Best Local Similarity 43.2%; Pred. No. 1.9e-26;
Matches 96; Conservative 30; Mismatches 89; Indels 7; Gaps 6;

QY 10 GVTSVQDEYRVQVQENNSNAOCLTINTATGAMTSTANFSGTGAPATPSIKGH 69
DB 46 GSTVYQ-GRYVQNNRMTSATQCVT--AIDSGRVTQAGOSVPTNKAPRSYSVNGCH 102
QY 70 WGNCTTKWVGPIQISQISAVTSMSTTOYSSGAYDAVDIMTNSPTTGGQNGEIM 129
DB 103 YTNC-SPTALPARISGSSAPSSISYGFVDNMYNNSYDIDLPTRTG-VNRTEIMI 160
QY 130 WLSRGVQPGSGRTATGTVVAGHTWVWQOQTSKTIISVLPATPSISNDLKAIFA 189
DB 161 WFNRRGOIPIQISQVGT-ASVAGRTWVWSSGNGINDVLSFV-APSAMSFSVDFMFEV 218
QY 190 DDAARGSLNSTDLDVNEAGFEIMOGGGLGSNSFSVTS 231
DB 219 ATVARGLAGNMTYLTISIQAGFEPQNGAGLAVNSFSVTVNG 260

RESULT 2
H70895

hypothetical protein RV1090 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence,revision 17-Jul-1998 #text,change 22-Oct-1999
C:Accession: J70895
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Seares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; M01D:98295987
A:Accession: H70895
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:AL021897; GB:AL123456; NID:93256022; PIDN:CA117206.1; PID:e125196
A:Experimental source: strain H37RV
C:Genetics:
C:Gene: RV1090

Query Match	18.9%	Score 235.5	DB 2	Length 151
Best Local Similarity	36.9%	Pred. No. 2.3e-11		
Matches	55	Conservative	21	Mismatches 70; Indels 3; Gaps 3
QY	80	MPIQISQISAVTSWSTQVSSGAYDAVDWNTNSTPTTGGPNCGETEIMTWINSRGVOP	139	
		::: :: :: ::		
Db	5	LPTVGGILSNAPTSIDYNYPTTGVADASIDICLDSPTPTTG-VNQGELMIWNNHGGSTOP	63	
QY	140	FGSQTATGVYAGHTWNNWOGQOOTSWKTIISYVLPNGATISMLDKAIFADAANGSLNT	199	
		::: :: :: :: :: :: :: :: :: :: :: :: :: :: ::		
Db	64	VGSPVGV-MTTIEGKNFVVMVDGNGGNMNAAYVATE-PIEVMSFDVMSFVDHATMEPTID	121	
QY	200	SDYLLDVEAGFEIMOGGOGISNSFSVSY	228	
		::: :: :: :: :: :: :: :: :: :: :: :: :: :: ::		
Db	122	SMYLTFSIRAGLEPMWSDGVGLGVDSFSAKY	150	

RESULT 3

S12610

cellulase (EC 3.2.1.4) precursor - *Aspergillus aculeatus*

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Aspergillus aculeatus*

C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 21-Jul-2000

C:Accession: S12610; S14118; S40186; JQ0458

R:Ooi, T.; Shimmyo, A.; Okada, H.; Murao, S.; Kawaguchi, T.; Arai, M.

Nucleic Acids Res. 18, 5884, 1990

A:Title: Complete nucleotide sequence of a gene coding for *Aspergillus aculeatus* cellulase

A:Reference number: S12610; MUID:91016934

A:Accession: S12610

A:Molecule type: DNA

A:Residues: 1-237 <OOI1>

A:Cross-references: EMBL:D00546; NID:9217818; PIDN:BAA00435.1; PID:9217819

R:Ooi, T.; Shimmyo, A.; Okada, H.; Hara, S.; Ikenaka, T.; Murao, S.; Arai, M.

Curr. Genet. 18, 217-222, 1990

A:Title: Cloning and sequence analysis of a cDNA for cellulase (FI-CMCase) from *Aspergillus*

A:Reference number: S14118; MUID:91064758

A:Accession: S14118

A:Molecule type: mRNA

A:Residues: 1-237 <OOI2>

A:Cross-references: EMBL:X52525; NID:92287; PIDN:CAA36757.1; PID:92288

A:Accession: S40186

A:Molecule type: protein

A:Residues: 17-18;42-49, 'X',51-54, 'X',66-79;90-111,136-205, 'XX',208-211 <OOI3>

C:Genetics:

A:Exons: 138/2; 212/1

A:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose

A:Pathway: cellulose degradation

C:Keywords: glycosidase; hydrolase; polysaccharide degradation; pyroglutamic acid

E:1-16/Domain: signal sequence #status predicted <SIG>

E:17-237/Product: cellulase #status experimental <KAT>

E:17/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental

```

Query Match      18.8%: Score 234; DB 2; Length 237;
Best Local Similarity 34.7%: Pred. No. 5, le-11;
Matches 76; Conservative 24; Mismatches 99; Indels 20; Gaps 10;

OY 17 GEYRVQINEMN---SSAQCCLTINTATGAMTV-SFA-NFSGGTGAPATYPYSIKGHWG 71
    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 31 GYVITNNMIMGKADSGSCCTIVNSASSAGTSWSTKRMNMGGENSEVKS-----YAN 81
    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
OY 72 NCTETKNVAMPIDISIGSASVTSMSTTOYSSGAYDVAVDIWTNSTPTTTGQPNGTEIMTL 131
    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 82 SGLTFNKRLIVDSIOI-PPTARMSYDNTGICRA-DVAIDLEFTAADINHWTMSDYELMTWL 139
    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
OY 132 NSRGGVOPFGSOTATGTVYAGHTVMNVWVGQQTSWKIIISVLTLPGATSISNLDKAFADA 191
    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 140 ARYGCVGPFIGSOIAT-AATWDGQTWEIWMYGANGSQSYTSFVAPTPIPSFG-DVNDFFKYL 197
    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
OY 192 AA-RGSLNTSDVLIDVEAGFEIWMGGOG-TGSNSVSAY 228
    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 198 TONGFPASSQYLLITLGTEPFTTGPPATILSNWSASY 236

```

```

RESULT      4
JU0328
Cellulase (EC 3.2.1.4) precursor - Erwinia carotovora subsp. carotovora
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase
C:Species: Erwinia carotovora subsp. carotovora
C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 08-Oct-1999
C:Accession: JU0328
R:Saarlahhti, H.T.; Henriksst, B.; Palva, E.T.
Gene 90, 9-14, 1990
A:Title: CelS: a novel endoglucanase identified from Erwinia carotovora subsp. carotovora
A:Reference number: JU0328; MUID:90337352
A:Accession: JU0328
A:Molecule type: DNA
A:Residues: 1-264 <SNAP>
A:Cross-references: GB:M32399; NID:g148389; PIDN:AAA24817.1; PID:g148390
A:Experimental source: strain SCC3193
A:Note: part of this sequence, including the amino end of the mature protein, was com
C:Genetics:
A:Gene: celS
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
E:1-32/Domin: signal sequence #status predicted <SIG>
F:3-264/Product: cellulase #status experimental <MAT>

```

	Query Match	17.3% Score 215; DB 2;	Length 264;
	Best Local Similarity	30.5%; Pred. No. 1,7e-09;	
	Matches 67;	Conservative 33;	Mismatches 90; Indels 30; Gaps
OY	16 DGEFVQVTNENSSA-----QQCLTINTATGA----WTVTANFGSGTGAPATYSIK	66	
Db	46 NKKYLLEFNVMGKDEIKMQOOTIFYNPSISGMWHMSPSTHSYK-----AYPLVS	97	
OY	67 GCCHGNCTTKNVGMPDIQSIGSAVTSWSTQVSSAGAYDAVDITWNTSTPTTGGQPNGT-	125	
Db	98 GMHTVAGTTEMSGLPIDQLSSSKMTSITSVNYIKATGYNNAAVDHFTTTDKAMDDSPTD	157	
OY	126 EIMTWLNSRGGVOPPGSOTATGTVTAAGTWNWOG-----QQTSWKLTISVYLPGATSIS	180	
Db	158 ELMTWLNDPT-NAGPAGDVIET-VFLDGSSNWVFEGMTINADNGGGMWNFSEVHTSGTNSAS	215	
OY	181 NLDKAIFADAAARGSLNTSD--YLLDPEAGFEIWMOGGO	218	
Db	216 -LNRIR-FTDYLVQTKQMMSDEKIISSEVEGETELF-GGDG	252	
RESULT	5		
	555931		

cellulase (EC 3.2.1.4) precursor - *Aspergillus niger*
 M:Alternate names: carboxymethylcellulase I; endo-1,4-beta-glucanase
 C:Species: *Aspergillus niger*
 C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
 C:Accession: S55931; S60657; J50730
 R:Sakamoto, S.; Tamura, G.; Ito, K.; Ishikawa, T.; Iwano, K.; Nishiyama, N.
 Curr. Genet. 27, 435-439, 1995
 A:Title: Cloning and sequencing of cellulase cDNA from *Aspergillus kawachii* and its expr
 A:Reference number: S55931; MUID:96059347
 A:Accession: S55931
 A:Molecule type: mRNA
 A:Residues: 1-239 <SAK>
 A:Cross-references: EMBL:DJ2901; NID:g217812; PID:g217813
 A>Note: the source is designated as *Aspergillus kawachii*
 A:Accession: S60657
 A:Molecule type: protein
 A:Residues: 76-86;176-186 <SAK2>
 A:Note: the source is designated as *Aspergillus kawachii*
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
 A:Pathway: cellulose degradation
 C:Keywords: blocked amino end; glycosidase; hydrolase; polysaccharide degradation
 F:1-16/Domain: signal sequence #status predicted <Sig>
 F:17-239/Product: cellulase #status predicted <Mat>

Query Match 14.8%; Score 184.5; DB 2; Length 239;
 Best Local Similarity 29.5%; Pred. No. 3.5e-07;
 Matches 67; Conservative 27; Mismatches 84; Indels 49; Gaps 13;

OY 13 SVQDEYRVQVNEW---NSSAOCLITN-TATGA-----WTVTANFSGTGAPATYP 62
 DB 25 SASSPYEVNOMLWEGYOGTSCVYDKLSSGASMTTMT-----WSGEG----- 72
 OY 63 STYKCHMGNCCTTK---NVGMPDQ---ISQISAVTS--WSTVSSGANDVADITFNS 114
 DB 73 -----TVKSYNSGLTFDKKLVSDVSIPTVMSODPTNOA-DVSYDLFTAA 120
 OY 115 TPTTGGPNGTEIMIMILNSRGVOPFGSQTATGYVAHNTNNWGGQOT---SMKITSY 170
 DB 121 NMDHATSSGDYELIMILWLRGSGVOPRIGQIAI-ATVGRKSMEWVIGTSTQGADEKITSF 179
 OY 171 VLTGATSTISNLDKAIKADAAA-RGSIMTSDYLLDVEAGEETWQCG 216
 DB 180 VAGSPINMSG-DIKDFNITLQNOGFRASQHLITLQCGTEPFTGG 225

RESULT 6
 G90360
 endoglucanase precursor [imported] - *Sulfolobus solfataricus*

C:Species: *Sulfolobus solfataricus*
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
 C:Accession: G90360
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, V.; Allard, G.; Aweyer, M.J.; Chan
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thl-Ngoc, H.P.; Redder, F
 arrett, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.
 submitted to Genbank, April 2001
 A:Description: *Sulfolobus solfataricus* complete genome.
 A:Reference number: A99139
 A:Accession: G90360
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-334 <KUR>
 A:Cross-references: GB:AE006641; NID:g13815224; PIDN:AAK42142.1; GSPDB:GN00155
 C:Genetics:
 A:Gene: SS01949

Query Match 10.7%; Score 133; DB 2; Length 334;
 Best Local Similarity 26.7%; Pred. No. 0.0048;
 Matches 71; Conservative 39; Mismatches 100; Indels 56; Gaps 18;

OY 3 TPGP---NONGVTSVQDEYRVQVNEWNS---SAAOCLITNTATGANTVS---TANFSG 52

DB 81 SPPIVYIANNSTVS-----FYLEVNMNNAKTWNGNTVTFNPLTTLTSLVSFNLQVNPLO 135
 OY 53 GTGAPATYPSIYKCGH-WGNCCTKAVGAPIQISQISAVTSW--STQVS-SGADVAY 108
 DB 136 WTNG-----YPEIYVGRKPMWTSYAGNI-FPMKRGNNTPFVWSYINLTKLDSINFDIAS 190
 OY 109 DIW-----TNSPTTGTQPNGTETIMILNSRGVOPFGSQTATGV-----TVAGHTWN 156
 DB 191 DAWIVRQIAFSPETANGNDIELMWLFSG-NLQPAQGVGEVVIPIYINHTLVNATFO 249
 OY 157 VMQOQTSWKLISYV-----LTPGATIS-NLDKAI-PADAARGSIN-TSDYLL 204
 DB 250 VMKKNVPMGWGMEYIAFRPDGMKVTNGYVAYEPMILFKALNFA-----SYNITNYLFT 303
 OY 205 DVEAGFEIMOGGGLGNSPSVSPTS 230
 DB 304 DWEFGTE-WGTMTSNGTAYFSWTISN 328

RESULT 7

S28634
 adhesin AIDA-I precursor - *Escherichia coli* plasmid pIB6
 C:Species: *Escherichia coli*
 C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 08-Oct-1999
 C:Accession: S28634; S22680; S28881; S72657
 R:Benz, I.
 submitted to the EMBL Data Library, March 1992
 A:Reference number: S28634
 A:Accession: S28634
 A:Molecule type: DNA
 A:Residues: 1-1286 <BEN>
 A:Cross-references: EMBL:X65022; NID:g42254; PIDN:CAA6156.1; PID:g42255
 R:Benz, I.; Schmidt, M.A.
 Mol. Microbiol. 6, 1539-1546, 1992
 A:Title: AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic *Esch*
 A:Reference number: S22680; MUID:92326638
 A:Accession: S22680
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 839-1286 <BE2>
 A:Cross-references: EMBL:X65022
 A:Experimental source: strain 2787
 A:Accession: S28881

A:Molecule type: protein
 A:Residues: 50-36 <BE3>
 A:Experimental source: strain 2787
 R:Suhr, M.; Benz, I.; Schmidt, M.A.
 Mol. Microbiol. 22, 31-42, 1996
 A:Title: Processing of the AIDA-I precursor: removal of AIDA and evidence for the out
 A:Reference number: S72657; MUID:97055419
 A:Accession: S72657
 A:Molecule type: protein
 A:Residues: 847-856 <SU>
 A:Experimental source: DAEC strain 2787
 C:Genetics:
 A:Genome: plasmid pIB6
 C:Keywords: membrane protein
 F:1-49/Domain: signal sequence #status predicted <Sig>
 F:50-1286/Product: adhesin AIDA-I #status predicted <Mat>

Query Match 10.4%; Score 129; DB 2; Length 1286;
 Best Local Similarity 25.9%; Pred. No. 0.045;
 Matches 69; Conservative 32; Mismatches 93; Indels 72; Gaps 13;

OY 7 NONGVTSVQDEYRVQVNEWNS---SAAOCLITNTATGANTVS---TANFSG----- 53
 DB 82 NSGGTQIVANNGGKTATTV-NSSGQ-----NVGTSGARITIVNSGGIQRVSSGVSASAT 136
 OY 54 --TGAPATY-----PSIYKCHMGNCCTKNNGM--PIQIS-----QISATVSMSTT 97
 DB 137 NLSGQNNIYNLGHASNFVIFSG---GNQITFSGGITDSTINISGCGQRVSSGVSASAT 193

```

OY 98 QVSSGAYDAVDIMTNSPTTTG-----QPNGTEIMIMLNSRGVOPFGS-QTATGVT 149
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 194 INSSQAQNLSEGAISITHSISGNOYISAGANATETIV---NSGFORVNSGAVATGTV 250
OY 150 VAGHTWVWVWOGQOTSMKIISYVLTFGATISNLDKAIADAA-----ARGSLN---- 198
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 251 LSGGTQVWVSSGSAI-----STSVYNSGVQTVAGATVDTTVNSGQNONISSG 299
OY 199 --TSDYLLDVEAGFEIWMGGGGLGSN 222
    ||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 300 GIVSETVWVSGTONIYSGSALSAN 325

RESULT 8
endoglucanase precursor [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: G90291
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aweez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.D.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: G90291
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-332 <KUR>
A:Cross-references: GB:AE006641; NID:g13814564; PIDN:AAK41590.1; GSPDB:GN00155
C:Genetics:
A:Gene: SSO1354

Query Match 10.3%; Score 128; DB 2; Length 332;
Best Local Similarity 25.0%; Pred. No. 0.012;
Matches 62; Conservative 38; Mismatches 98; Indels 50; Gaps 14;

OY 19 YAVQTNEMSSAQO-----CLITNTAGAMTVSTANFSGGTGAPATPSIYKGC 68
    ||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 93 FYLEVMMMAKYNNGVTVFNPPLARTLSVFNLTQVKLEWTVNG-----YPEIYVGR 145
OY 69 H-WGNOTKTVGMPIQISQSAVTSW--STQVVS-SGAYDAVDIM-----TNSPTTT 119
    ||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 146 KWDPAVYAGNT-FPMKIGMMPFVWSFYTNLTLPDSINFDIASDAWIVPOIATSPGTA 204
OY 120 GQNGTEIMIMLNSRGVOPFGSQTATGV-----TVAGHTWVWVWOGQOTSMKIISYV- 171
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 205 PGNQIEIWMVLFSG-NIQPAGEQVGVVPIYINHTLVNAIFQWEMKSVRWGMEYIA 263
OY 172 -----LTPGATIS-NIDKAIADAAAGSLN-TSDYLLDVEAGFEIWMGGGGLGSN 222
    ||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 264 FRPDQMKVTVNGVSYEPNLFKAL-----SNFTSYNTITNYLLDMEFGTE-MGTMTSNGTA 318
OY 223 SPSVATVS 230
    ||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 319 YFSWTVSN 326

RESULT 9
polymorphic membrane protein E/F family TC0262 [imported] - Chlamydia muridarum (straln
C:Species: Chlamydia muridarum, Chlamydia trachomatis MOpn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: G81722
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MOpn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255
A:Accession: G81722
A:Status: preliminary

```

```

A:Molecule type: DNA
A:Residues: 1-1025 <TET>
A:Cross-references: GB:AE002293; GB:AE002160; NID:g7190298; PIDN:AAF39131.1; PID:g719
A:Experimental source: strain Nigg (MOpn)
C:Genetics:
A:Gene: TC0262

Query Match 9.0%; Score 112; DB 2; Length 1025;
Best Local Similarity 22.6%; Pred. No. 0.72;
Matches 61; Conservative 25; Mismatches 94; Indels 90; Gaps 12;

OY 7 NONGVTSVODGERVQTNEMSSAQOCLITNTAGATVSTAN-----FEGGTGGAPA 59
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 536 NQGTAP-----RIRVN-----PQDMQNTNSQAPVSTENVATQKIFGSLVDE 584
OY 60 TYPPIYKCGHMGNCCTTKNMGPIQISQSAVTSWSTQVSSGAYDAV---DIWTNST 115
    ||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 585 NYESVYDSCDLLSRGANOPILHETTNDAQLSNDKNT-LNTSLYSLEPHYGYGLMTSNW 643
OY 116 PTTT---GQNGTEIMIMLWV-----SRGVOPFGSOTA 145
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 644 MTTTFTVSLNSTETQTNANNSIQEOKNTSETPDSNSTTAKIPSIASGTGTPLAT--- 700
OY 146 TGVTVAGHTWVWVWOGQOTSMKIISYVLTFGATISNLDKAIADAAAGSL----- 197
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 701 TDVTVTRHSLV-----SWPPIGYI-----ADPARGDLIANNLVSS 737
OY 198 --NTSDYLLDVEAGFEIWMGGGGLGSNSFS 225
    ||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 738 GRNTTLYRLSLLPD-DSWALQGSATLFT 766

RESULT 10
hypothetical protein H02F09.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2000
C:Accession: T33369
R:Gelsel, C.; Harmon, G.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid H02F09.
A:Reference number: 221330
A:Accession: T33369
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1275 <GEI>
A:Cross-references: EMBL:AF077538; PIDN:AA064622.1; GSPDB:GN00028; CESP:H02F09.3
C:Genetics:
A:Gene: CESP:H02F09.3
A:Map position: X
A:Introns: 42/1; 78/3; 106/3; 135/2; 161/2; 224/1
C:Superfamily: Yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida

Query Match 8.7%; Score 108; DB 2; Length 1275;
Best Local Similarity 23.5%; Pred. No. 1.9;
Matches 59; Conservative 33; Mismatches 103; Indels 56; Gaps 9;

OY 3 TPGPNQVTSVODGERVQTNEMSSAQOCLITNTAGATV---STANFSGGTG--- 55
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 755 TPGTESTITT---DGSVSGSTGSGTSTNNPGSTDSSTGISTVSGSSLSITSGSGTV 811
OY 56 -----GAPATPSIYKCGHMGNCCTTKNMGPIQISQSAVTSWSTQVSSG 102
    ||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 812 SGSSDMTVSTGSSPSGSESTVSG---ASTMSPISGSSVEFTSGSSVSTVSGSTSSS- 867
OY 103 AYDAVDIMTNSPTTTGQNGTEIMI-WLNSRGV-QPFGSQTATGTVVAGHTWVWOG 160
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 868 -----TTGQSTVSESVTSVSESTISQSTGTTGEGSVFSGSTSTANG 912
OY 161 QOTSMKIISYVLTFGATISNLDKAIADAAAGSLNTSDYLLDVEAGFEIWMGGGGLG 220

```

Db	913	SSITMASATGSHDFPGST-----ESTITGIVTGESEIVSGSGTGTITREGSHISE	960
		:	
QY	221	SNSEFSVSVTSIG 221	
		: :	
Db	961	STMHTTVGVSHTG 971	

```

RESULT 11
T34369
hypothetical protein T19D12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34369

```

A;INTRONS: 36/1; 134/2; 180/1; 622/3; 691/2; 754/1; 1111/2; 1174/1; 1271/3; 1322/2; 1681

RESULT 12
H72240
endoglucanase - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: H72240
R:Nelson, R.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.U.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316
A:Accession: H72240
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-258 <ARN>
A:Cross-references: GB:AE001800; GB:AE000512; NID:g4962090; PIDN:AAD36591.1; PID:g498205
A:Experimental source: strain MSB8
C:Genetics:
C:Gene: TML524

[illegible]

RESULT 13
AB2583
conserved hypothetical protein XF2237 [Imported] - *Xylella fastidiosa* (strain 9a5c)
C:Species: *Xylella fastidiosa*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: AB2583
R:Anonymous: The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A:Reference number: AB2515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: AB2583
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1..958 <SIM>
A:Cross-references: GB:AE004036; GB:AE003849; NID:9107384; PID:AAF85036.1; GSPDB:GN
R:Experimental source: strain 9a5c
R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Brlones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
as-Melo, E.; Docena, C.; El-Dorri, H.; Fachinani, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kempner, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurame, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsushima, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF2237

[illegible]

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:38:45 ; Search time 38.44 Seconds
(Without alignments)
232.680 Million cell updates/sec

Title: US-09-917-384-7
Perfect score: 1244
Sequence: 1 DCTPGPNNGVTYSVODGEYR.....IMCGGGLGNSFSVSTSG 231

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	234	18.8	237	1 GUN_ASPAC	P22669 aspergillus
2	215	17.3	264	1 GUN_ERMCA	P16630 erwina car
3	184.5	14.8	239	1 GUN_ASPAC	Q12679 aspergillus
4	129	10.4	1286	1 AIDA_ECOLI	Q03155 escherichia
5	112	9.0	1025	1 PMP_CHLMU	O9P146 chlamydia m
6	105	8.4	1036	1 HP12_DEIRA	P13136 delinococcus
7	104	8.4	747	1 GUN_CELFI	P50400 cellulomona
8	98	7.9	3075	1 LMAI_HUMAN	P25381 homo sapien
9	96.5	7.8	948	1 HP11_DEIRA	P56867 delinococcus
10	96.5	7.7	1076	1 HP11_DEIRA	P20676 saccharomyc
11	96	7.7	479	1 CB8A_SULSO	P58029 sulfolobus
12	96	7.7	1475	1 N153_HUMAN	P49790 homo sapien
13	95	7.6	211	1 YOR3_SOUV3	P00450 southampton
14	95	7.6	666	1 MUR2_ENTHR	P39046 enterococcu
15	94.5	7.6	1251	1 YQ03_CABEL	P09550 caenorhabdi
16	94	7.6	657	1 CSPI_CORGL	Q01377 corynebacte
17	93.5	7.5	380	1 PG11_SCLSC	Q12708 sclerotinia
18	93	7.5	335	1 XYNB_STRLI	P26515 streptomyce
19	92.5	7.4	863	1 YEJO_ECOLI	P33924 escherichia
20	92	7.4	313	1 KREL_YEAST	P17260 saccharomyc
21	91	7.3	197	1 XYNA_SCHCO	P35809 schizophyil
22	91	7.3	525	1 NU62_RAT	P17955 rattus norv
23	91	7.3	550	1 MYCO_STRCI	P20910 streptomyce
24	90.5	7.3	2124	1 PGCA_RAP	P07897 rattus norv
25	89.5	7.2	796	1 YSBA_CABEL	P09625 caenorhabdi
26	89	7.2	610	1 CHIT_STRLI	P11220 streptomyce
27	89	7.2	718	1 CDGT_BACCI	P309458 bacillus c
28	88.5	7.1	697	1 CRAC_DICDI	P35401 dictyostell
29	88	7.1	225	1 XYN2_ASPNG	P55330 aspergillus
30	88	7.1	430	1 NU42_YEAST	P49666 saccharomyc
31	88	7.1	598	1 YAU6_SCHPO	Q10168 schizosacch
32	88	7.1	718	1 CDGT_BACCS	P31747 bacillus sp
33	87.5	7.0	159	1 MP63_MYCTU	P97175 mycobacteri

34	87.5	7.0	530	1 Y9A_SCHPO	Q09768 schizosacch
35	87.5	7.0	773	1 CDH_PHACH	Q01738 phanerocha
36	87.5	7.0	887	1 QVAI_DROME	Q23969 drosophila
37	87	7.0	225	1 XYNL_EMENT	P55332 emeticella
38	87	7.0	605	1 YHCB_YEAST	P38739 saccharomyc
39	87	7.0	1447	1 VGL2_CVPT	Q01977 porcine tra
40	87	7.0	1449	1 VGL2_CVPT	P18450 porcine tra
41	86.5	7.0	713	1 CDG2_PABMA	P31835 penicillium
42	86.5	7.0	966	1 FIB1_PETMA	P02674 petromyzon
43	86.5	7.0	1025	1 SIAP_CAUCR	P35828 caulobacter
44	86	6.9	886	1 VGP3_EBYA8	Q07284 epstein-bar
45	86	6.9	1045	1 GUNB_CELFI	P26225 cellulomona

ALIGNMENTS

RESULT	1	ALIGNMENTS
GUN_ASPAC	STANDARD;	PRT; 237 AA.
AC P22669;		
DT 01-AUG-1991 (Rel. 19, Created)		
DT 01-AUG-1991 (Rel. 19, Last sequence update)		
DT 15-DEC-1998 (Rel. 37, Last annotation update)		
DE Endoglucanase I precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)		
DE (Cellulase) (FI-CMCase).		
OS Aspergillus aculeatus.		
OC Eukaryotes; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;		
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.		
OX NCBI_TaxID=5053;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=F-50;		
RX MEDLINE=91016934; PubMed=2216782;		
RA Ooi T., Shimmyo A., Okada H., Murao S., Kawaguchi T., Arai M.;		
RT "Complete nucleotide sequence of a gene coding for Aspergillus		
aculeatus cellulase (FI-CMCase)."		
RL Nucleic Acids Res. 18:5884-5884(1990).		
RN [2]		
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.		
RC STRAIN=F-50;		
RX MEDLINE=91064758; PubMed=2249253;		
RA Ooi T., Shimmyo A., Okada H., Hara S., Ikenaka T., Murao S.;		
RA Arai M.;		
RT "Cloning and sequence analysis of a cDNA for cellulase (FI-CMCase)		
from Aspergillus aculeatus".		
RL Curr. Genet. 18:217-222(1990).		
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic		
linkages in cellulose.		
CC -1- SUBCELLULAR LOCATION: Secreted.		
CC -1- INDUCTION: BY CELLULOSIC MATERIALS AND HEMICELLULOSES.		
CC -1- MISCELLANEOUS: WILL ALSO HYDROLYSE 1,4-LINKAGES IN BETA-D-GLUCANS		
ALSO CONTAINING 1,3-LINKAGES.		
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY H (FAMILY 12 OF GLYCOSYL		
HYDROLASES).		
CC This SWISS-PROT entry is copyright. It is produced through a collaboration		
between the Swiss Institute of Bioinformatics and the EMBL outstation -		
the European Bioinformatics Institute. There are no restrictions on its		
use by non-profit institutions as long as its content is in no way		
modified and this statement is not removed. Usage by and for commercial		
entities requires a license agreement (See http://www.isb-sib.ch/announce/		
or send an email to license@isb-sib.ch).		
CC EMBL; D00546; BA00435.1; -		
DR EMBL; X52525; CA36757.1; -		
DR PIR; JQ0458; JQ0458.		
DR PIR; S12610; S12610.		
DR PIR; S14118; S14118.		
DR Interpro: IPR002594; Glyco_hydro_12.		
DR Pfam: PF01670; Glyco_hydro_12; 1.		
DR ProDom: PD004316; Glyco_hydro_12; 1.		
KW Cellulose degradation; Hydrolyase; Glycosidase; signal.		


```

Db 25 SASSPPSYNONIMGEYQCTGSCVTVYDKLSSGSAWHTKW-----WSGEG----- 72
QY 63 STYKGCCHMCNCTTK---NYGMPIQ---ISOIGSAVTS--WSTOVSSGAYDAVADITWNS 114
Db 73 -----TVKSYNSNGSLFPDKRLVDVSIPTSVWMSODDTWVQA-DVSYDLFTAA 120
QY 115 PTTTGGPNCETIMLWLNRSRGVQPGSOTATGTVAGHTWVWGQOQT---SMKIISY 170
Db 121 NADHATSSGDIYELMIAWLGSGVQPIGKQIAT-AIVGKSMWEVWYGTSTQAGAEKQTSF 179
QY 171 VLTGATSTSNIDKAIKAFDAAA-RSLMTSDLLDVEAGFEIMOG 216
Db 180 VAGSPINWSG-DIKDFNYLTQNGFPASSOHLITLQCEPTEPTG 225

RESULT 4
AIDA_ECOLI STANDARD; PRT: 1286 AA.
AC 003135:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 01-JUN-1994 (Rel. 29, Last annotation update)
DE Adhesin aida-1 precursor.
GN AIDA-1.
OS Escherichia coli.
OG Plaemid p186.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.
RC STRAIN=2787 (O126:H27);
RX MEDLINE=92326638; PubMed=1625582;
RA Benz I., Schmidt M.A.;
RT "AIDA-1, the adhesin involved in diffuse adherence of the
RT diarrhoeagenic Escherichia coli strain 2787 (O126:H27), is
RT synthesized via a precursor molecule."
RL Mol. Microbiol. 6:1539-1546(1992).
CC -1- FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE
CC ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI
CC TO EPITHELIAL CELLS.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X65022; CAA46156.1; -.
DR PIR: S28634; S28634.
KW Cell adhesion; Signal; Outer membrane; Plasmid.
FT SIGNAL 1 49
FT CHAIN 50 ? ADHESIN AIDA-1.
FT PROPEP ? 1286
SO SEQUENCE 1286 AA; 132271 MW; B2A00F72AC05FB34 CRC64;

Query Match 10.4%; Score 129; DB 1; Length 1286;
Best Local Similarity 25.9%; Pred. No. 0.011;
Matches 69; Conservative 32; Mismatches 93; Indels 72; Gaps 13;

QY 7 NONGCVSYODGERVQTNEMNSSAOCLTINTATGATVSTANFSG----- 53
Db 82 NSGGTQIVNNGGKTTATTV-NSSGSO-----NVTSGATISTIVNSGCIORVSSGVAASAT 136
QY 54 --TGAPARY-----PSYKGCCHMCNCTTKVGM--PIQIS-----OIGSAVTSMTST 97
Db 137 NLSGGMNINYNLGHASNYVIFSG--GNQITFGSGITDSTINISSGGOQRSSGCVASNTT 193
QY 98 QVSSGAYDAVADITWNSPTTTC-----QPNGTEIMLWLNRSRGVQPGS--QTATGVT 149

```

```

Db 194 INSSGAMNILEGAIETHISSGNGNOISAGANATERIV---NSGGFORVNSGAVATGV 250
QY 150 VAGHTWVWGQOQTSMKIISYVLTGATSTSNIDKAIKAFDAA-----ARGLN----- 198
Db 251 LSGGTQVNSGGSAT-----STVYNSGVQTFAGATVTTFTVNSGNNISSG 299
QY 199 --TSDYLLDVEAGFEIMOGGGLGNS 222
Db 300 GIVSETTVANSGTQNIYSGGSALSAN 325

RESULT 5
PMPF_CHLMU STANDARD; PRT: 1025 AA.
ID PMPF_CHLMU
AC 09PL46:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pmpF precursor (Polymorphic membrane
DE protein F).
GN PMPF OR FC0262.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mopn / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Ueteyback T., Berry K., Bass S.,
RA Linher K., Weldman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE002293; AAF39131.1; -.
DR TIGR: FC0262;
DR InterPro: IPR003357; OMP.
DR Pfam: PF02385; OMP; 1.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 1025 POTENTIAL.
SO SEQUENCE 1025 AA; 11458 MW; EA472E07818B9PF2 CMC64;

Query Match 9.0%; Score 112; DB 1; Length 1025;
Best Local Similarity 22.6%; Pred. No. 0.19;
Matches 61; Conservative 25; Mismatches 94; Indels 90; Gaps 12;

QY 7 NONGCVSYODGERVQTNEMNSSAOCLTINTATGATVSTAN-----FSGGTGAPA 59
Db 536 NDGGTTAP-----RIRVN-----PQDMTQNTNSNOAPVSTENVAQKIFSGSLVSLVDE 564
QY 60 TYSYIKGCCHMCNCTTKVGMPIQISGSAVTSMTTQVSSGAYDAV---DIPTNST 115
Db 585 NYESVYDSCDLSGKANKNPILHETITNDQLSDMKRT-LNTSLYSLPHRYGGLWTSNW 643
QY 116 PTTT---GQPNTEIMLWLN-----SRGGVQPGFSQTA 145

```

Query Match	8.4%	Score 105	DB 1	Length 1036
Best Local Similarity	24.9%	Pred. No. 0.68		
Db	644	MTTFEVLVLTNSTETQGTANNSSIQEQKNTSEFDSNSTTAAIPSRISTAGGTPLAT----	700	
Oy	146	TCGYVAGHTNANWVGQOQTSKMTIISVLTLPGATSIISNLDLKAIPDAARGL-----	197	
Db	701	TDYVYTRHSLVY-----SWTPIGYI-----ADPARRGDLIANNLVSS	737	
Oy	198	--NTSDYLDVEAGFEIMWGOGGLGSNSFS	225	
Db	738	GRNTLYLRSLLPD-DSWFLAQSAATVLT	766	
RESULT	6			
ID	HP12_DEIRA	STANDARD	PRT: 1036 AA.	
AC	P13126			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Hexagonally packed intermediate-layer surface protein precursor.			
GN	HPI.			
OS	Deinococcus radiodurans.			
OX	Bacteria; Thermus/deinococcus group; Deinococcales; Deinococcus.			
NCBI	TaxID=1299;			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	STRAIN-SARK;			
RC	MEDLINE=86032846; PubMed=3667529;			
RT	Peters J., Peters M., Lotzspeich F., Schaefer W., Baumeister W.;			
RT	"Nucleotide sequence analysis of the gene encoding the Deinococcus			
RT	radiodurans surface protein, derived amino acid sequence, and			
RT	complementary protein chemical studies.";			
RL	J. Bacteriol. 169:5216-5223(1987).			
CC	-1- FUNCTION: SHAPE MAINTENANCE, POSSIBLE PROTECTION FROM NOXIOUS			
CC	ENZYMES OR EXOGENOUS AND UNSETTLING DNA, AND MAY MEDIATE HOMOTYPIC			
CC	CELL-CELL CONTACTS.			
CC	-1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A			
CC	HEXAGONAL S-LAYER.			
CC	-1- CONTAINS TIGHTLY BOUND REDUCING SUGARS (SIX PER POLYPEPTIDE			
CC	CHAIN) AND FATTY ACIDS (COVALENTLY BOUND AND LOCATED IN THE N-			
CC	TERMINAL REGION).			
CC	-1- PPM: THE N-TERMINUS IS BLOCKED.			
CC	-1- MISCELLANEOUS: THE HYDROPHILIC C-TERMINAL REGION RICH IN AROMATIC			
CC	AA COULD BE ENGAGED IN INTERACTIONS WITH NUCLEIC ACIDS, AND THE			
CC	BOUND FATTY ACIDS AND THE N-TERMINAL REGION COULD SERVE TO ANCHOR			
CC	THE LAYER TO THE OUTER MEMBRANE OF D.RADIOURANS. HPI LAYER			
CC	CONTAIN ABOUT 30% BETA STRUCTURE AND VIRTUALLY NO ALPHA HELIX.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.jsb-sib.ch/announce/			
CC	or send an email to license@lsb-sib.ch).			
CC	-----			
DR	EMBL; M17895; AAA23335.1; -			
DR	PIR; A29832; A29832.			
DR	HSSP; P11795; 2TBV.			
DR	PROSITE; PS00013; PROKAR_LIPOPROTEIN; UNKNOWN_1.			
KW	Glycoprotein; Lipoprotein; Signal; Cell wall; S-layer.			
FT	SIGNAL	1	17	POTENTIAL.
FT	CHAIN	18	1036	HEXAGONALLY PACKED INTERMEDIATE-LAYER
FT				SURFACE PROTEIN.
FT	DISULFID	74	86	PROBABLE.
FT	DISULFID	256	275	
FT	DISULFID	642	754	
FT	DOMAIN	60	250	SER/THR-RICH (29%).
FT	DOMAIN	985	1030	RICH IN AROMATIC AA (29%).
SO	SEQUENCE	1036 AA;	108028 MW;	AAFF98206A74AEEE CRC64;

	Matches	62:	Conservative	32:	Mismatches	111:	Indels	44:	Gaps	11:
QY	10	GVTG----	VODGEKRVQGTENMNSSAOQCLTINTAGTAVSTANESGGTG-----	55						
		::	::	::	::	::	::	::	::	
Dd	144	GVSSAPITIKRDANGNVVGYDNTVNDNATILTVARGVTVTAGNVSGNGPFTTRFDLS	203							
QY	56	GAPATPSIPIYKGCHMKCTTKNVG-MPQLQISGSATVSMSITVOVSSCAIVDAIDI--WT	112							
		::	::	::	::	::	::	::	::	
Dd	204	GGNDQVTLTANYQA-GTTTPPVGSINILITPAVGTSVSGSFTVRVT---FKRANEVCMV	258							
QY	113	NSTPTTGDPNGTE---IMVLNSRGVPGCSOTATGVTAAGHTWMMWGQQTSMKTIIS	169							
		::	::	::	::	::	::	::	::	
Dd	259	GGAALVTAQVDVSTSYSCDYVPVNSTGNV----VITMGKGVNGQIVTATRNI SVTQAAS	314							
QY	170	Y-VLTPGATSISMDLKAFADA-----AARGLMNTSDYLIDVEAGEFIWOGGGLGSN	222							
		::	::	::	::	::	::	::	::	
Dd	315	YGVAVTPADGDQLTLTSEIVRADPDAGWRRLGGVSTSPDPLMLNDIYIK-----CTV	366							
QY	223	SFSYSVTSG 231								
		::	::	::	::	::	::	::	::	
Dd	367	NFSVANAPAG 375								
		::	::	::	::	::	::	::	::	
RESULT	7									
GUND_CELFT	ID	GUND_CELFT	STANDARD:	PRT:	747	AA.				
AC	P50400;									
Dt	01-OCT-1996	(Rel. 34, Created)								
Dt	01-OCT-1996	(Rel. 34, Last sequence update)								
Dt	01-OCT-1996	(Rel. 34, Last annotation update)								
DE	Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)									
GN	CEND.									
OS	Cellulomonas fimi.									
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;									
CC	Actinomycetales; Micrococccineae; Cellulomonadaceae; Cellulomonas.									
OX	KCB1_TaxID=1708;									
RN	[1]									
RP	SEQUENCE FROM N.A.									
RX	MEDLINE=93209933; PubMed=8458833;									
RA	Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.;									
RT	"Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase									
RL	D (Cend)', a family A beta-1,4-glucanase.";									
-1	J. Bacteriol. 175:1910-1918(1993).									
CC	-1 CATALYTIC ACTIVITY: Endohydrolaysis of 1,4-beta-D-glucosidic									
CC	linkages in cellulose.									
CC	-1 PATHWAY: CELLULOSE DEGRADATION.									
CC	-1 SIMILARITY: CONTAINS 2 FIBROECTIN TYPE III-LIKE DOMAINS.									
CC	-1 SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN									
CC	(CBD).									
CC	-1 SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL									
CC	HYDROLASES).									
CC	-----									
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration									
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -									
CC	the European Bioinformatics Institute. There are no restrictions on its									
CC	use by non-profit institutions as long as its content is in no way									
CC	modified and this statement is not removed. Usage by and for commercial									
CC	entities requires a license agreement (See http://www.isb.ch/announce/									
CC	or send an email to license@lsb.slb.ch).									
CC	-----									
DR	HMBL; L02544; AAA3089.1; .									
DR	HSSB; P07986; IEXG.									
DR	InterPro: IPR001919; CBD_2.									
DR	InterPro: IPR003961; FN.ITI.									
DR	InterPro: IPR003962; FN.III.repeat.									


```

FT DOMAIN 1150 1159 LAMININ EGF-LIKE 14 (N-TERMINAL).
FT DOMAIN 1160 1361 LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT DOMAIN 1162 1402 LAMININ EGF-LIKE 14 (C-TERMINAL).
FT DOMAIN 1403 1451 LAMININ EGF-LIKE 15.
FT DOMAIN 1452 1508 LAMININ EGF-LIKE 16.
FT DOMAIN 1509 1555 LAMININ EGF-LIKE 17.
FT DOMAIN 1556 2116 DOMAIN II AND I.
FT DOMAIN 2117 2297 LAMININ G-LIKE 1.
FT DOMAIN 2305 2481 LAMININ G-LIKE 2.
FT DOMAIN 2486 2673 LAMININ G-LIKE 3.
FT DOMAIN 2713 2885 LAMININ G-LIKE 4.
FT DOMAIN 3070 2890 LAMININ G-LIKE 5.
FT DOMAIN 1706 1795 COILED COIL (POTENTIAL).
FT DOMAIN 1968 2088 COILED COIL (POTENTIAL).
FT DOMAIN 2088 2120 COILED COIL (POTENTIAL).
FT SITE 2534 2536 CELL ATTACHMENT SITE.
FT DISULFID 270 279 BY SIMILARITY.
FT DISULFID 272 290 BY SIMILARITY.
FT DISULFID 292 301 BY SIMILARITY.
FT DISULFID 297 305 POTENTIAL.
FT DISULFID 304 324 BY SIMILARITY.
FT DISULFID 327 336 BY SIMILARITY.
FT DISULFID 329 361 BY SIMILARITY.
FT DISULFID 364 373 BY SIMILARITY.
FT DISULFID 376 394 BY SIMILARITY.
FT DISULFID 397 409 BY SIMILARITY.
FT DISULFID 399 427 BY SIMILARITY.
FT DISULFID 429 438 BY SIMILARITY.
FT DISULFID 441 451 BY SIMILARITY.
FT DISULFID 454 467 BY SIMILARITY.
FT DISULFID 466 471 BY SIMILARITY.
FT DISULFID 473 482 BY SIMILARITY.
FT DISULFID 485 500 BY SIMILARITY.
FT DISULFID 742 751 BY SIMILARITY.
FT DISULFID 744 757 BY SIMILARITY.
FT DISULFID 760 769 BY SIMILARITY.
FT DISULFID 772 788 BY SIMILARITY.
FT DISULFID 791 806 BY SIMILARITY.
FT DISULFID 793 816 BY SIMILARITY.
FT DISULFID 819 828 BY SIMILARITY.
FT DISULFID 831 846 BY SIMILARITY.
FT DISULFID 849 863 BY SIMILARITY.
FT DISULFID 851 870 BY SIMILARITY.
FT DISULFID 873 882 BY SIMILARITY.
FT DISULFID 885 899 BY SIMILARITY.
FT DISULFID 902 914 BY SIMILARITY.
FT DISULFID 904 921 BY SIMILARITY.
FT DISULFID 923 932 BY SIMILARITY.
FT DISULFID 935 948 BY SIMILARITY.
FT DISULFID 951 963 BY SIMILARITY.
FT DISULFID 953 969 BY SIMILARITY.
FT DISULFID 971 980 BY SIMILARITY.
FT DISULFID 983 995 BY SIMILARITY.
FT DISULFID 998 1007 BY SIMILARITY.
FT DISULFID 1000 1014 BY SIMILARITY.
FT DISULFID 1016 1025 BY SIMILARITY.
FT DISULFID 1028 1041 BY SIMILARITY.
FT DISULFID 1044 1056 BY SIMILARITY.
FT DISULFID 1065 1074 BY SIMILARITY.
FT DISULFID 1074 1087 BY SIMILARITY.
FT DISULFID 1077 1142 BY SIMILARITY.
FT DISULFID 1403 1412 BY SIMILARITY.
FT DISULFID 1405 1419 BY SIMILARITY.
FT DISULFID 1422 1431 BY SIMILARITY.
FT DISULFID 1434 1449 BY SIMILARITY.
FT DISULFID 1452 1466 BY SIMILARITY.
FT DISULFID 1454 1476 BY SIMILARITY.
FT DISULFID 1479 1488 BY SIMILARITY.
FT DISULFID 1491 1506 BY SIMILARITY.
FT DISULFID 1509 1521 BY SIMILARITY.
FT DISULFID 1511 1528 BY SIMILARITY.
FT DISULFID 1530 1539 BY SIMILARITY.
FT DISULFID 1542 1553 BY SIMILARITY.

```

```

FT DISULFID 1556 1556 INTERCHAIN (PROBABLE).
FT DISULFID 1560 1560 INTERCHAIN (PROBABLE).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 555 555 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 763 763 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 952 952 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1045 1045 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1407 1407 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1579 1579 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1596 1596 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1678 1678 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1689 1689 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1698 1698 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1717 1717 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1804 1804 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1894 1894 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 7.9%; Score 98; DB 1; Length 3075;
Best Local Similarity 22.1%; Pred. No. 7.9;
Matches 66; Conservative 40; Mismatches 108; Indels 84; Gaps 15;

QY 9 NGVTSVQGEYRVQNMENS-----SAQOCLTINTATGATV----- 45
DB 2184 SSTRLEPPDPFIDNRWHSIHVAPNGISLSYKEMSNQSPKTSKSPETAVLDVN 2243
QY 46 -STANFSGGTG---APATPSIYKCHWNC---TTKVGAPRISOIGSAVTSMTQ 98
DB 2244 NSTLNFVGGIGQIKKSPAVKVTHERKGC-LGRAFLNGSIGMWNTERGKRCGFGSSQ 2302
QY 99 VSSGAYDV---AYDIWNSTPTTQO-----PNGTEIMWINS-----R 134
DB 2303 NEDPSFHEDSGSYVEKSLPATVTOIIMLFNTFSPNG--LILYGSYKTDLSIEFR 2360
QY 135 GGVO---PPGSGTAVGVYAGHTNVW-----QG---QGTSMKIIISVILP 174
DB 2361 GVKWMTDLGSGPTILLDRRNNGTWYKIAFORNRKGVLAIVDAVYNSNETKOGETP 2420
QY 175 GATG-ISMILKAIPA-----DMAARGSLNTSDYLLDVEAGFEIMOGGGLGNSFSV 226
DB 2421 GASDLNLRDKRPPIYVGGLPRSRVVRGVTTKSPFGCIK-NLEISRSFDDLNRNGV 2477

RESULT 9
HP1L DEIRA STANDARD; PRT; 948 AA.
AC P56867;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hexagonally packed intermediate-layer surface protein precursor.
GN HPI OR DR2508.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarewicz K.S., Atavand L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.,
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RI."
RL Science 286:1571-1577(1999).
CC -!- FUNCTION: SHAPE MAINTENANCE, POSSIBLE PROTECTION FROM NOXIOUS
CC ENZYMES OR EXOGENOUS AND UNSETTLING DNA, AND MAY MEDIATE HOMOTYPIC
CC CELL-CELL CONTACTS (BY SIMILARITY).

```


16-OCT-2001 (Rel. 40, last sequence update)
 01-MAR-2002 (Rel. 41, last annotation update)
 Cytochrome b558/566 subunit A.
 CESA OR SSO2801 OR C48_011.
 Sulfolobus solfataricus.
 Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 NCBI_TaxID=2287;
 [1]
 SEQUENCE FROM N.A.
 STRAIN-ATCC 35092 / DSM 1617 / P2;
 MEDLINE=21332296; PubMed=11427726;
 She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 Aveyer M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
 De Moers A., Erasmo G., Fletcher C., Gordon P.M.K.,
 Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
 "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
 Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 -1- FUNCTION: MONOHEME CYTOCHROME WHOSE PHYSIOLOGICAL FUNCTION IS NOT
 YET CLEAR (BY SIMILARITY).
 -1- SUBCELLULAR LOCATION: Integral membrane protein.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

 DR EMBL: AE006874; AAK2914.1.
 KW Electron transport; Transmembrane; Heme; Glycoprotein;
 KW Complete proteome.
 FT TRANSMEM 19 39 POTENTIAL.
 FT TRANSMEM 456 476 POTENTIAL.
 FT CARBOHYD 73 73 N-LINKED (POTENTIAL).
 FT CARBOHYD 99 99 N-LINKED (POTENTIAL).
 FT CARBOHYD 152 152 N-LINKED (POTENTIAL).
 FT CARBOHYD 172 172 N-LINKED (POTENTIAL).
 FT CARBOHYD 182 182 N-LINKED (POTENTIAL).
 FT CARBOHYD 191 191 N-LINKED (POTENTIAL).
 FT CARBOHYD 219 219 N-LINKED (POTENTIAL).
 FT CARBOHYD 262 262 N-LINKED (POTENTIAL).
 FT CARBOHYD 288 288 N-LINKED (POTENTIAL).
 FT CARBOHYD 302 302 N-LINKED (POTENTIAL).
 FT CARBOHYD 325 325 N-LINKED (POTENTIAL).
 FT CARBOHYD 348 348 N-LINKED (POTENTIAL).
 FT CARBOHYD 385 385 N-LINKED (POTENTIAL).
 SQ SEQUENCE 479 AA; 52532 MW; 7C92B0B91A8245B4 CRC64;
 Query Match 7.7%; Score 96; DB 1; Length 479;
 Best Local Similarity 24.9%; Pred. No. 1.5; Mismatches 79; Indels 76; Gaps 14;
 Matches 58; Conservative 20;
 4 PGPNGNGTSTVODGERVRO---TNEWSSAQOCLTINTAGTAVTSTANFSGGTGGAPA 59
 DB 252 PQAFDFGTFVNDGSLVPGGANIM-----MW-VSGATWNN-----A 289
 QY 60 TTPSTYIKGCHMGNCCTKNVGNP-IQISQISAVTSKSTVOSSGAVD-AYIWTNSTPT 117
 DB 290 TYDPAKFWLWON--TSLTGLPYIDPDNNGFAVPLYTN--WTNNYEVDVTAGIM--YTPV 342
 QY 118 TTPGNGTFTIMTWSRGVGPFGSOTAGTVAGTVMWMOGQOTSMKTIISYVLTGPAF 177
 DB 343 TTSGLNGSLFTW-----TGATYONGKWTV-----EFAPLAVPSA- 378
 QY 178 SISNDLKAIFAADAAAGSLNTSYDLLDVEAGFEIWOOGGGLGNSFSVSYS 230
 DB 379 -----YAKWMPNITVGTYYV-----AFAYWOG--KIGETLFDKSTIS 414

RESULT 12
 N153_HUMAN
 ID N153_HUMAN STANDARD: PRT; 1475 AA.
 AC P49790;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Nuclear pore complex protein Nup153 (Nucleoporin Nup153) (153 kDa
 nucleoporin).
 GN NUP153.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94154002; PubMed=8110839;
 RA McMorris I., Bastos R., Horton H., Burke B.;
 "Sequence analysis of a cDNA encoding a human nuclear pore complex
 protein, hnup153.";
 RT Biochim. Biophys. Acta 1217:219-223(1994).
 RL -1- FUNCTION: POSSIBLE DNA-BINDING SUBUNIT OF THE NUCLEAR PORE
 COMPLEX (NPC). THE REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN
 ANCHORING COMPONENTS OF THE PORE COMPLEX TO THE PORE MEMBRANE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX. LOCATED TO THE
 CC -1- TERMINAL RING STRUCTURE OF THE NUCLEOPLASMIC CAGE.
 CC -1- DOMAIN: CONTAINS X-F-X-F-X-F-G REPEATS.
 CC -1- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
 CC SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
 CC NUP1, NSP1, POM 121 AND MAMMALIAN P62.
 CC -1- SIMILARITY: CONTAINS 4 RANBP2-TYPE ZINC FINGERS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: Z25535; CAA80982.1; -
 DR MIM: 603948;
 DR InterPro; IPR001876; Znf-RanBP.
 DR Pfam; PF00641; Zf-RanBP; 4.
 DR SMART; SM00547; Znf_RBZ; 4.
 DR PROSITE; PS01358; ZF_RANBP2_1; 4.
 DR PROSITE; PS0199; ZF_RANBP2_2; 4.
 KW Nuclear protein; Transport; Repeat; Zinc-finger; DNA-binding.
 FT DOMAIN 4 14
 FT DOMAIN 443 447 POLY-GLY.
 FT ZN_FING 657 687 RANBP2-TYPE 1.
 FT ZN_FING 722 751 RANBP2-TYPE 2.
 FT ZN_FING 793 822 RANBP2-TYPE 3.
 FT ZN_FING 851 880 RANBP2-TYPE 4.
 SQ SEQUENCE 1475 AA; 153889 MW; 3CB415A6909DE80E CRC64;
 Query Match 7.7%; Score 96; DB 1; Length 1475;
 Best Local Similarity 27.5%; Pred. No. 5.1; Mismatches 47; Indels 30; Gaps 7;
 Matches 38; Conservative 23;
 41 GAMVSTANFSGGTGAPATYPSIYKCH-----WGNCCTKNVGNPFIQISQIG 88
 DB 1342 GSISSTALFPTGSPAPPTFGTVSSSSQPYFGGQPSGASRGSGTTPSSSAF---QFG 1398
 QY 89 SAVTSNSTQVS-SCAYDYAVIDWTNSPTTTGPGNGEFTIMTWSRGVOPFGSQTATG 147
 DB 1399 SSTTNFTNNSPSGVFTFGAN---SSTPAASAOPSGS-----GGF-PF-NQSPAA 1444
 QY 148 VTVAGTVMWMOGQOTSW 165
 DB 1445 FTVGSGKKNVSSSGTSF 1462

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:37:51 ; Search time 125.36 Seconds
(without alignments)
318.777 Million cell updates/sec

Title: US-09-917-384-7
1244
Sequence: 1 DCTPGPNQNGVTSVQDEYR.....IMGGGIGGNSFSVSTSG 231

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp_Archea:*
2: sp_Bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_Organella:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	472.5	38.0	371	2	09KIH1 streptomyc
2	449.5	36.1	377	2	008468 streptomyc
3	436.5	35.1	382	2	059963 streptomyc
4	434.5	34.9	384	2	09X602 streptomyc
5	425.5	34.2	381	2	09R3J3 streptomyc
6	424.5	34.1	381	2	054331 streptomyc
7	407.5	32.8	260	2	033897 rhodotherm
8	235.5	18.9	151	16	053438 mycobacteri
9	214.5	17.2	264	2	031030 pectobacter
10	208	16.7	263	2	09AN41 bradyrhizob
11	207.5	16.7	244	2	09XYS4 streptomyc
12	193.5	15.6	239	3	013454 aspergillus
13	192.5	15.5	239	3	074705 aspergillus
14	175.5	14.1	234	3	000095 trichoderma
15	175.5	14.1	238	3	094218 aspergillus
16	150	12.1	239	3	09P8N6 cochliobol

17	133	10.7	334	17	097X08 sulfolobus
18	128	10.3	332	17	097Y67 sulfolobus
19	109	8.8	402	10	09AXB0 oryza sativ
20	108	8.7	407	3	000895 colletoxic
21	108	8.7	1275	5	076602 caenorhabd1
22	108	8.7	1844	5	022579 caenorhabd1
23	107.5	8.6	319	1	09V2T0 pyrococcus
24	107	8.6	392	10	096487 lycopersico
25	106	8.5	780	2	054229 streptomyc
26	105.5	8.5	257	2	060032 thermotoga
27	105.5	8.5	258	16	09S5X8 thermotoga
28	105.5	8.5	958	16	09PBA7 xylella las
29	105	8.4	461	2	059922 streptomyc
30	100.5	8.1	464	9	09A286 lactobacill
31	100.5	8.1	1203	5	09N5K0 caenorhabd1
32	100	8.0	4776	16	097P71 streptococc
33	99.5	8.0	267	15	09WDJ9 human immun
34	99.5	8.0	267	15	09WDK6 human immun
35	99	8.0	267	15	09WEA0 human immun
36	99	8.0	335	2	09ROB8 cellulomona
37	99	8.0	427	2	09FCO7 pseudomonas
38	98.5	7.9	267	15	09W9X0 human immun
39	98.5	7.9	526	11	063850 mus sp. nuc
40	98.5	7.9	3930	16	09B220 rhizobium
41	98	7.9	795	10	09LJ20 oryza sativ
42	98	7.9	1390	5	017602 caenorhabd1
43	98	7.9	1749	16	P73032 synechocyst
44	97.5	7.8	1160	11	09WUN1 mus musculu
45	97.5	7.8	1198	11	099PM6 mus musculu

ALIGNMENTS

RESULT 1
ID 09KIH1 PRELIMINARY; PRT: 371 AA.
AC 09KIH1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLULOSE 12A.
GN CEL12A.
OS Streptomyces sp. 11A68.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=133452;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=11A68;
RA van Solingen P., Meljer D., van der Kleij W.A.H., Barnett C.C.,
RA Holle R., Power S.D., Jones B.E.;
RT "Cloning and expression of an endocellulase gene from a novel
RT Streptomycete isolated from an East African soda lake."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF233376; AAF91283.1; -
DR HSSP: P07986; 1EXG.
DR InterPro: IPR002594; Glyco_hydro_12.
DR InterPro: IPR001230; Prenyltn.
DR Pfam: PF01670; Glyco_hydro_12; 1.
DR ProDom: PD004316; Glyco_hydro_12; 1.
DR PROSITE: PS00294; PRENYLATION; UNKNOWN_1.
SQ SEQUENCE 371 AA; 38481 MW; 0E1BC4288A148914 CRC64;

Query Match 38.0%; Score 472.5; DB 2; Length 371;
Best Local Similarity 46.6%; Pred. No. 3.4e-27;
Matches 104; Conservative 28; Mismatches 82; Indels 9; Gaps 8;

QY 10 GVSIVQGEYRVQVTEMNSSAQCCLTINTAGATVSTANSSGTCGAPATPSIYIGCH 69
DB 40 GTTITD-RVYVQNNRNGTSATOC--INTVGNGEFTIQADGSVPTNGAPRSPSYVYDGH 96

OY	70	WGNCTTKVNHGDIQISQIGSAVTSMSTNOVSSGVADVADIDMTNPTTGGPNNGEIMI	129
Db	97	YGNCAPRTT-LPMKISSIGSAPSSVSTRYTGTGVTNAVDIWDPPRTNG-VNRTEIMI	154
OY	130	WLNSRGVOPFGSOTATGTVTAHTMNWMOGOOTSMMKIISYVLTPGATSINMLDKAIPA	189
Db	155	WFNRVGVPVPIGSPVGT-AHVCGRSMEWWTGSNGSNDVISF-LAPSALISWSFDVKD-FV	211
OY	190	DAAARGSLNTSD-YLLDYBAGFEITMOGGOGGLGSNSFSVSYSIG	231
Db	212	DOAVSHGLATPDWYLTLTSIOAGEPMEWGGTGLAVNSFFSSAVNMG	254
RESULT	2		
ID	008468	PRELIMINARY:	PRT: 377 AA.
AC	008468:		
DT	01-JUL-1997	(TREMBLrel. 04, Created)	
DT	01-JUL-1997	(TREMBLrel. 04, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE	Cel2 (EC 3.2.1.4).		
DN	CelA2.		
OS	Streptomyces halstedii.		
OC	Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;		
CC	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.		
OX	NCBI_TaxID=1944;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-JM8.		
RX	MEDLINE-97307849; PubMed-9182697;		
RA	Garcia-Salas A.L., Fernandez-Abalos J.M., Sanchez P., Ruiz-Atribas A.,		
RT	Santamaria-Sanchez R.I.;		
RT	"Two genes encoding an endoglucanase and a cellulose-binding protein		
RT	are clustered and co-regulated by a TTA codon in Streptomyces		
RL	halstedii JM8.";		
RJ	Biochem. J. 324:403-411(1997).		
DR	EMBL; U51222; AAC45429.1; -.		
DR	HSSP; P07986; IEXG.		
DR	InterPro; IPR001919; CBD_2.		
DR	InterPro; IPR002594; Glyco_hydro_12.		
DR	Pfam; PF00553; CBD_2; 1.		
DR	Pfam; PF01670; Glyco_hydro_12; 1.		
KW	Prodrom; PD004316; Glyco_hydro_12; 1.		
KW	Hydrolase; Glycosidase.		
SO	SEQUENCE 377 AA; 38829 MW; 9892191B741EC280 CRC64;		
Query Match	36.1%; Score 449.5; DB 2; Length 377;		
Best Local Similarity	43.8%; Pred. No. 1.7e-25;		
Matches 98; Conservative 31; Mismatches 88; Indels 7; Gaps			
OY	8	ONGVTSVODGEERVOJTNEMNSSAOOCLINTATGAMTVSTANFSGTGGAAPTSPYIKG	67
Db	44	QYGSTTIQ-GRVYVONNNKGASAPCVT--AIDSGFRYQAAGCAVPTNKAPKSPSVFMG	100
OY	68	CHWGCTTKRNKMPLQIQISOIGSAVTSWSTTOVSSGAYDVADIVMTNSTPTTGGPNGTEI	127
Db	101	CHYTNC-SPGTLMIPAOVSGIASAPSSISYGFVGSAYVNASTYAIWDLPTRKNG-VNRTEI	158
OY	128	MIMLNRSBGVOPFGSOTATGTVTAHTMNWMOGOOTSMMKIISYVLTPGATSINMLDKAI	187
Db	159	MIMLNKRVPIPIGSIQAGT-ASVGGRTWMVMGNSGNSNDVISFV-APSAVASMSFDVMD	216
OY	188	FADAARGLSLNTSDYLLDYBAGFEITMOGGOGGLGSNSFSVSYSIG	231
Db	217	VNRITARGMAQNMYLTLSYQACFEFWONGAGILAVNSFSSTVNLG	260
RESULT	3		
ID	059963	PRELIMINARY:	PRT: 382 AA.
AC	059963:		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	

```

DT      01-NOV-1996 (TREMBLrel_01, last sequence update)
DT      01-DEC-2001 (TREMBLrel_19, last annotation update)
DE      CELLULOSE [EC 3.2.1.4].
GN       EGLS.
OS      Streptomyces rochei (Streptomyces parvullus).
OC      Bacteria; Firmicutes; Actinobacterlia; Actinobacteridae;
CC      Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
CX      NCBI_TaxID=1928;
RN       [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=A2.
RX      MEDLINE=95011642; PubMed=7523249;
RA      Perito B., Hanhart E., Iranil T., Iqbal M., McCarthy A.J.,
R          Mastromiel G.;
RT      "Characterization and sequence analysis of a Streptomyces rochei A2."
RL      Gene 148:119-124(1994).
DR      EMBL: X73953; CAAS2139.1; -.
DR      HSSP: P07986; IEXG.
DR      InterPro: IPRO01919; CBD_2.
DR      InterPro: IPRO02594; Glyco_hydro_12.
DR      Pfam: PF00553; CBD_2; 1.
DR      Pfam: PF01670; Glyco_hydro_12; 1.
DR      Prodom: PD004316; Glyco_hydro_12; 1.
DR      PROSITE: PS00561; CBD_BACTERIAL; UNKNOWN_1.
KW      Hydrolyase; Glycosidase.
SQ      SEQUENCE   382 AA; 39398 MW; 21C014342EFC6565 CRC64;

Query Match           35.1% Score 436.5; DB 2; Length 382;
Best Local Similarity 43.2%; Pred. No.1.ee-24;
Matches    96; Conservative 30; Mismatches 89; Indels 7; Gaps

QY      10 GVTSVDGEYNVQNTNEMNSSAQOCLTITAGTAATVTANFSGGTGGAPATPSITYKCH 69
        | :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db       46 GSYYIQ-GRYVVQNRRNGSTATQCVT--AIDSGRTRYQADGSVPTNKAPSISPVFGCH 102
        : || : :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY      70 WNCCTTKNVGPDIQSIGSAVTSMTTOVSOGAYDVAYDWTNSTPTTTQQPNGTEIMI 129
        : || : :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db       103 YTNCG-SPTALPARISGISSPASSISYGFDNAVNAASYDWIPLDPTRPDG-VNRTEIMI 160
        : || : :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY      130 WLNSRGGVPPRSQTATGCATVTAAGHTWVMWGQQTSMKIISVLTPGATSSINDLKAIFA 168
        | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db       161 WFNRVGQIQPIGSOVGT-ASVAGRFTWEWSGNGTFNDVLSFY-APSAMSSMSFDVMDFVR 218
        ||| ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY      190 DAAARGSLINTSDYLDDVEAGEFIKGGGGLSGNSFSYSVTSG 231
        ||| ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db       219 ATVARGLAGNDMYLTSLIOAGFEPPNQMGAGLAIVNSFSSTVNTG 260
        ||| ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

RESULT         4
Q9X602          PRELIMINARY;             PRT;            384 AA.
AC      O9X602;
DT      01-NOV-1999 (TREEBLrel_12, Created)
DT      01-NOV-1999 (TREEBLrel_12, Last sequence update)
DT      01-JUN-2001 (TREEBLrel_17, Last annotation update)
DE      CELLULOSE.
GN       CELSI.
OS      Streptomyces viridosporus.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC      Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX      NCBI_TaxID=67581;
RN       [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=T7A;
RA      Ramachandran S., Crawford D.L.;
RT      "Characterization and sequence analysis of two genes involved in
RT      cellulase degradation in Streptomyces viridosporus T7A, and its
RT      expression in Escherichia coli".
RL      Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR      EMBL: AF130408; AAD25090.1; -.
DR      HSSP: P07986; IEXG.
DR      InterPro: IPRO01919; CBD_2.
```

DR InterPro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR ProDom: PD004316; Glyco_hydro_12; 1.
 DR PROSITE: PS00561; CBD_BACTERIAL; UNKNOWN_1.
 DR SEQUENCE 384 AA; 40918 MW; D3968B6E8DEDE65 CRC64;

Query Match 34.9%; Score 434.5; DB 2; Length 384;
 Best Local Similarity 42.7%; Pred. No. 2, 2e-24;
 Matches 102; Conservative 28; Mismatches 90; Indels 19; Gaps 9;

QY 2 CTRGPNQGVTSVQDEYRQTMENSSAQCCLINTATGAMVSTANFSGGTGAPATY 61
 DB 42 CEP-----YGTFTIQ-GRVYQNNRNGSSPOCYTA-TDTG-FRLTQADGVPVNGAPKSY 94
 QY 62 PSYKNCNMGNCNT--TK-----NVGMPFQISQISGAVTSWSTTOVSSGAYDAVDIWT 112
 DB 95 PSYKNCNMGNCNTPGKRLPARISGISAPARIGISSAPSSISYGVGCAVYNASDIWL 134
 QY 113 NSTPTTGGPNGTEIMIMWLSRGGVQPGSGQTATGVTAGHTWNVWOGQOTSMKIIISYVL 172
 DB 155 DPTPRPDG-VNRREIMWFKVGPDIQISQVGT-ATVGRTWQVNSGGNSNDVLSFV- 211
 QY 173 TREATISNLDKAFADAAARSLNTSDYLLDVEAGFEIMOGGCGIGNSFSVSTSG 231
 DB 212 APSAIESWSEFVDVDFVETVARGMAQNDWYLTISVQAGFEPMONGAGLAVNSFSSTVDTG 270

RESULT 5

Q9RJY3 PRELIMINARY; PRT; 381 AA.

AC 09RJY3; 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PUTATIVE SECRETED CELLULOSE B.
 GN CELB.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Cereno A.M., Parkhill J., Bartell B.G., Rajandream M.A.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE-97000351; PubMed-8843436;
 RA Redenbach M., Kleier H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL133210; CAB61599.1; -;
 DR HSSP: P07986; 1EXG.
 DR InterPro: IPR001919; CBD_2.
 DR InterPro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR ProDom: PD004316; Glyco_hydro_12; 1.
 DR PROSITE: PS00561; CBD_BACTERIAL; UNKNOWN_1.
 DR SEQUENCE 381 AA; 39199 MW; 97CD8FE58679E4EC CRC64;

Query Match 34.2%; Score 425.5; DB 2; Length 381;
 Best Local Similarity 42.3%; Pred. No. 1e-23;

Matches 94; Conservative 32; Mismatches 89; Indels 7; Gaps 7;

QY 10 GYTSVQDEYRQTMENSSAQCCLINTATGAMVSTANFSGGTGAPATYPSYKNC 69
 DB 49 GTTITIQ-GRVYQNNRNGSRATQCYTA-TDTG-FRVTQADGSAPTNAPSPSVFNCGH 105
 QY 70 WGNCTTKNVMPFQISQISGAVTSWSTTOVSSGAYDAVDIWTNSTPTTGGPNGTEIMI 129
 DB 106 YTNC-SPTGALPVRALDVTSAAPSISYGFVDGAVYNASYDIWLPARTDG-VNQTETMI 163
 QY 130 WLSRGGVQPGSGQTATGAVTAGHTWNVWOGQOTSMKIIISYVLPLGATISNLDKAFIFA 189
 DB 164 WFNRRVGPDIQISQVGT-ASVGRTWQVNSGGNSNDVLSFV-APSAISGSEFVDVDFV 221
 QY 190 DAARGSLNTSDYLLDVEAGFEIMOGGCGIGNSFSVSTSG 231
 DB 222 ATYARGLAENDWYLTISVQAGFEPMONGAGLAVNSFSSTVETG 263

RESULT 6

Q54331 PRELIMINARY; PRT; 381 AA.

AC 054331; 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CELLULOSE B PRECURSOR.
 GN CELB.
 OS Streptomyces lividans.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1916;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-66;
 RA Wiltmann S., Shareck F., Kluepfel D., Morosoli R.;
 RT "Purification and characterization of the CelB endoglucanase from
 Streptomyces lividans 66 and DNA sequence of the encoding gene."
 RL Appl. Environ. Microbiol. 60:1701-1703(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-66;
 RA Shareck F.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U04629; AAB71950.1; -;
 DR HSSP: P07986; 1EXG.
 DR InterPro: IPR001919; CBD_2.
 DR InterPro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR ProDom: PD004316; Glyco_hydro_12; 1.
 DR PROSITE: PS00561; CBD_BACTERIAL; UNKNOWN_1.
 KW SIGNAL.
 FT SIGNAL 1 40 POTENTIAL.
 FT CHAIN 41 381 CELLULOSE B.
 DR SEQUENCE 381 AA; 39239 MW; A7E99BF590A24EC CRC64;

Query Match 34.1%; Score 424.5; DB 2; Length 381;
 Best Local Similarity 42.3%; Pred. No. 1, 2e-23;
 Matches 94; Conservative 32; Mismatches 89; Indels 7; Gaps 7;

QY 10 GYTSVQDEYRQTMENSSAQCCLINTATGAMVSTANFSGGTGAPATYPSYKNC 69
 DB 49 GTTITIQ-GRVYQNNRNGSRATQCYTA-TDTG-FRVTQADGSAPTNAPSPSVFNCGH 105
 QY 70 WGNCTTKNVMPFQISQISGAVTSWSTTOVSSGAYDAVDIWTNSTPTTGGPNGTEIMI 129
 DB 106 YTNC-SPTGALPVRALDVTSAAPSISYGFVDGAVYNASYDIWLPARTDG-VNQTETMI 163
 QY 130 WLSRGGVQPGSGQTATGAVTAGHTWNVWOGQOTSMKIIISYVLPLGATISNLDKAFIFA 189

	Matches	66:	Conservative	38:	Mismatches	98:	Indels	29:	Gaps	11:
OY	15	ODGEKRVQTNEMNSA-----OQCILINATG-----WTVSTANSGCGAGAPATPSTY	65							
Db	45	ENNRKYEVNNWGWKGDEVKGMDOQTVEYNPSPTSGMWMHMPSSSVK-----AVPSLV	96							
OY	66	KCHHGNCCTTKVNPAPIOISQISAFTSWSTQVSSGA-YDAVAVDIWNSTPTTGOPNG	124							
Db	97	SCWHMTAAVTENSGLPIKLSS-NKSTITSNVYSISTATLAAAYDVMFHTTIDAKMSDRP	155							
OY	125	T-ETMIWLNSRCGVOPFCSSGTATGVTAAGHWNVMO-----GGQTSWKIIIVLPGATS	178							
Db	156	TDELIMTWLNNT-MACPADDYEIT-VFLDSDSSNNVFKEGWINAONGGMMNFSEVRITSNINS	213							
OY	179	ISNDLKAIIPADAAARGS-LNTSDYLVLVEAGEFELIMOGGQGIGSNSFSVSU	228							
Db	214	AS-LNIIRFPNTDVGQRKKMWMAKIISSEVLGETEIFGDCQIDITIKWSYDV	263							
RESULT	10									
09ANM1	ID	PRELIMINARY;	PRT;	263	AA.					
09ANM1	ID									
DT	01-JUN-2001	(TREMBLrel. 17, Created)								
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)								
DT	01-OCT-2001	(TREMBLrel. 18, Last annotation update)								
DE	10568.									
GN	10568.									
OS	Bradyrhizobium japonicum.									
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;									
OX	Bradyrhizobium group; Bradyrhizobium.									
RN	NCBI_Taxid=375;									
RN	[1]									
RP	SEQUENCE FROM N.A.									
RC	STRAIN=110SPC4;									
RX	MEDLINE=21101824; PubMed=11157954;									
RA	Gottfert M., Rothlieberger S., Kundig C., Beck C., Marty R.,									
RA	Hennecke H.;									
RT	"Potential symbiosis-specific genes uncovered by sequencing a 410-kb									
RL	DNA region of the Bradyrhizobium japonicum chromosome.";									
RL	J. Bacteriol. 183:1405-1412(2001).									
DR	EMBL; AF322013; AAG60937.1;									
DR	InterPro: IPR002594; Glyco_hydro.12.									
DR	Prodom: PD004316; Glyco_hydro.12.1.									
DQ	SEQUENCE 263 AA; 28509 MW; 8f66BAFe939487d6 CRC64;									

	Query Match	16.7%	Score 208;	DB 2;	Length 263;	
	Best Local Similarity	28.2%;	Pred. No. 6.1e-08;			
	Matches	69;	Conservative	43;	Mismatches	75; Indels 58; Gaps 14.
Oy	12 TSVODGEYRVQNTENWS-----SAQQCLITTFAT--GAMTVSTANFSGGTGAPATYPS	63	:	:	:	:
Db	50 SSAOYGSSSIDGYSMNNNVWVGKAGAPQTIVSAYVNWGMWSNP-----DTGGI-KSPY-	102	:	:	:	:
Oy	64 IYKGCHMNCCTTKNYGNPDIQSISAYTSWBTTOVSSGAYVDYAVDWTNRPPTTGQN	123	:	:	:	:
Db	103 -----HEG----FNVGKP--LSTITFLISNFQEVPTSGAMDVAADIMOS-----N	143	:	:	:	:
Oy	124 GTETIMIMLN-----SRGVOPFGSQTA-----TGVTYAGHTMNVMQGQTSWKII	168	:	:	:	:
Db	144 QYEIMLMNTNYNGPNPDGGGVAKPISYRYKASSGPALPIYTIVNDVYGATWNVEFG-ENNHHXYI	202	:	:	:	:
Oy	169 SYVLTPGATSTISNLDKAIFADPAAR---GSJNTSDYLIDVAGEFTIWOGGGLCSNSFS	225	:	:	:	:
Db	203 SFELR-SKTNSGTVDIKSKYLVLOWIKSKGYFGEINVG---NWQYGEVETISSPGGMENFEFN	257	:	:	:	:
Oy	226 SVSTS	230	:	:	:	:
Db	258 WTITS	262	:	:	:	:
RESULT	11					

09KYS4			
ID	09KYS4	PRELIMINARY;	PRT; 244 AA.
AC	09KYS4;		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	PURATIVE SECRETED SUGAR HYDROLASE.		
GN	SC5H4.15.		
OS	Streptomyces coelicolor.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
OX	NCBI_TaxId:1902;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-A3(2);		
RC	Murphy L., Harris D.;		
RA	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-A3(2);		
RC	Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;		
RA	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.		
RL	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-A3(2);		
RC	MEDLINE-97000351; PubMed-8843436;		
RA	Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,		
RA	Klasni H., Hopwood D.A.;		
RT	"A set of ordered cosmids and a detailed genetic and physical map for		
RT	the 8 Mb Streptomyces coelicolor A3(2) chromosome.";		
RL	Mol. Microbiol. 21:77-96(1996).		
DR	EMBL; AL355913; CAB91127.1;		
DR	InterPro; IPR002594; Glyco_hydro.12.		
DR	Pfam; PF01670; Glyco_hydro.12.1.		
DR	ProDom; PD004316; Glyco_hydro.12.1.		
DR	Hydrolase.		
SQ	SEQUENCE 244 AA; 26123 MW; 2CA65242798EAD4 CRC64;		

Query Match	Similarity	16.7%	Score	207.5	DB 2	Length	244
Best Local	Similarity	26.3%	Prod. No.	6.1e-08			
Matches	65	Conservative	44	Mismatches	84	Indels	37
						Gaps	12

ID	Accession	Species	Length	Score	DB
QY	1	NONGVTSVODGEHYRVQVNEWNSA-QOCLTINTAT-GAMTVSTANFSGGTGAAPATYPS	63		
DB	47	DQWGMTSLNG--YTLVNNIMGSGAGSCQCVMANSGTDMGVW-ADHPPTDG-----IKSYN	98		
QY	64	IYKGCWNCCTTKKNGMPQIQISQISAVTSWSTTOYSSAIVYANDIYTNSTPTTGGQPN	123		
DB	99	-----ANKVINKP--ITSLSLSTSYNVTVPSSCAVMTSDYIMPTD-----Y	138		
QY	124	GEIMIMWLSRSGVOPFGSQGTATGTVAGHTNNWQOQTSKIIISYVLPGATS--ISN	181		
DB	139	DYELMWNNHNGAVPPLTGFQGS-VGLGHTIDYTKKNGANAEVRSFLTSPNSNGTVNI	197		
QY	182	LDLKAIFADAAARGSLNTSDYLLDVEAGEEIMQGGGIGLGSNSFSVSYTSG	231		
DB	198	LPIIKWID--TKGWMG-NETIGDVQFGEIITSAGGLDFPRNNLTVSGG	244		

RESULT	12
ID	013454
AC	013454
DT	01-JAN-1998 (Tremblrel. 05. Created)
DT	01-JAN-1998 (Tremblrel. 05. Last sequence update)
DT	01-DEC-2001 (Tremblrel. 19. Last annotation update)
DE	ENDO-1.4-BETA-GLUCANASE (EC 3.2.1.4).
GN	CELA.
OS	Aspergillus oryzae.
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC	Eurotiaceae; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX	NCBI_TaxID=5062;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-KBN616;
 RA MEDLINE-97161783; PubMed-9008887;
 RA Kitamoto N., Go M., Shibayama T., Kimura T., Kito Y., Ohmura K.,
 RA Tsukagoshi N.;
 RT "Molecular cloning, purification and characterization of two endo-1,4-
 RT beta-glucanases from *Aspergillus oryzae* KBN616."
 RL Appl. Microbiol. Biotechnol. 46:538-544(1996).
 DR Interpro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR Prodom: PD004316; Glyco_hydro_12; 1.
 KW Hydrolase; Glycosidase.
 SQ SEQUENCE 239 AA; 26096 MW; C0F850E5DPEB455D CRC64;

 Query Match 15.6%; Score 193.5; DB 3; Length 239;
 Best Local Similarity 27.4%; Pred. No. 6.3e-07;
 Matches 60; Conservative 39; Mismatches 87; Indels 33; Gaps 10;

 QY 13 SVODGEYRVQTNW--NSSAOOCLTIN--TATGAMTVSTANFSGTGAPATYPSIYNG 67
 DB 25 SASPPYSVNNMLMGODSGCYVDNLSSGAAMHTTWTMNGEGSVKS-----76
 QY 68 CHMGCTCKNMGMPQIOISGSAVTSMTQVSSGAYDAVDIWTNSTPTTGGPNGTEI 127
 DB 77 -YSNATVFFDKLVSDVOSIPTDV--EWSODNTNVA-DVAYDLFTAAADONHTVYSGDYEL 133
 QY 128 MIMLSRGGVOPFGSQTATGTVAGHTMNVWOG-----QOTSMKIIISYVLPATIS 180
 DB 134 MIMLARIGTIOPIGTOIDT-ATVEGHTWELWGTITIGAGAEKRTSVS-----ATPILN 186
 QY 181 NL--DLKAFADAAARGSLNLS-DYLLDVEAGFEIWMOG 216
 DB 187 TFGDIKKEFFDYITSKHSFPASAQYLINMQFTPEPTGG 225

 RESULT 13
 ID 074705 PRELIMINARY; PRT; 239 AA.
 AC 074705;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ENDOGLUCANASE A PRECURSOR (EC 3.2.1.4) (CELLULOSE)
 DE (ENDO-1,4-BETA-GLUCANASE A) (CARBOXYMETHYLCCELLULASE).
 GN EGIA OR CEKA.
 OS *Aspergillus niger*.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; *Aspergillus*.
 OX NCBI_TaxID=5061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CBS 120.49 / N400;
 RX MEDLINE-98432774; PubMed-9758775;
 RA van Peijl N.M.M.E., Gietkens M.M.C., de Vries R.P., Visser J.,
 RA De Graaf L.H.;
 RT "The transcriptional activator XlnR regulates both xylinolytic and
 RT endoglucanase gene expression in *Aspergillus niger*."
 RL Appl. Environ. Microbiol. 64:3615-3619(1998).
 CC -1- FUNCTION: HAS CARBOXYMETHYLCCELLULASE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY H (FAMILY 12 OF GLYCOSYL
 CC HYDROLASES).
 DR EMBL: AJ224451; CA11964.1; -
 DR Interpro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR Prodom: PD004316; Glyco_hydro_12; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; signal.
 FT SIGNAL 1 16
 FT CHAIN 17 239 ENDODGLUCANASE A.

SQ SEQUENCE 239 AA; 25859 MW; 5EB6267DA54AA2A CRC64;

 Query Match 15.5%; Score 192.5; DB 3; Length 239;
 Best Local Similarity 28.8%; Pred. No. 7.5e-07;
 Matches 63; Conservative 33; Mismatches 90; Indels 33; Gaps 11;

 QY 13 SVODGEYRVQTNW--NSSAOOCLTIN--TATGA----WTVSTANFSGTGAPATYP 62
 DB 25 SASPPYSVNNMLMGEOYGTGSCYVDNLSSGASMTHEWT-----MGCGEGTVKS---76
 QY 63 SYKGGHMGCTCKNMGMPQIOISGSAVTSMTQVSSGAYDAVDIWTNSTPTTGGP 122
 DB 77 -----YSNATVFFDKLVSDVOSIPTSV--EMKODNTNVA-DVAYDLFTAAAVDHATSS 128
 QY 123 NGTEIMIMLSRGVOPFGSQTATGTVAGHTMNVWOGQOT-----SMKIIISYVLPATIS 178
 DB 129 GYELMIMLARIGTIOPIKQIAT-ATVGGKMEWYSGTTQAGAEQRITSVSESPINS 187
 QY 179 ISNLDKAFADAAA-RGSLNLSDYLLDVEAGFEIWMOG 216
 DB 188 YSG-DINAFPSYLTQNGCPASSQYLINQFTEAFTGG 225

 RESULT 14
 ID 000095 PRELIMINARY; PRT; 234 AA.
 AC 000095;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ENDO-BETA-1,4-GLUCANASE (EC 3.2.1.4).
 GN EGL.
 OS *Trichoderma reesei* (Hypocrea jecorina).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Hypocreaceae; *Hypocrea*.
 OX NCBI_TaxID=51453;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OM9414;
 RA Okada H., Tada K., Sekiya T., Yokoyama K., Takahashi A., Tonda H.,
 RA Kumagai H., Morikawa Y.;
 RT "Molecular characterization and heterologous expression of the gene
 RT encoding a low-molecular-mass endoglucanase from *Trichoderma reesei*
 RT OM9414."
 RL Appl. Environ. Microbiol. 64:55-563(1998).
 DR EMBL: AB003694; BAA20140.1; -
 DR Interpro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR Prodom: PD004316; Glyco_hydro_12; 1.
 KW Hydrolase; Glycosidase.
 SQ SEQUENCE 234 AA; 25159 MW; DF476EDE384AD1 CRC64;

 Query Match 14.1%; Score 175.5; DB 3; Length 234;
 Best Local Similarity 26.7%; Pred. No. 1.3e-05;
 Matches 60; Conservative 34; Mismatches 94; Indels 37; Gaps 9;

 QY 19 YRVQTNWENSSAOOCLTINTATGAMTVSTANFSGTGAPATYPSIYKCHMGCTCKNV 78
 DB 31 YTVSNMIMKASA-----GSGFGCVTAVLSGSA-----SMHADWMSG--GQNN 72
 QY 79 GMPQIOISQIG-----SAVTSMTQVSSGAY-----DVAYDIWTNSTPTTGGPNG 124
 DB 73 VRSYQNSOIAIPQKRTVNSISSMPTT--ASWSYSGSNIRANAVYDLFTAAANHTVYSGD 130
 QY 125 TEIMIMLSRGVOPFGSQTATGTVAGHTMNVWOGQOTSMKIIISYVLPATISINDL 184
 DB 131 YELMIMLARIGTIOPIGIGSSQGT-VVNGGQSWTLXYGNCAMQVYSPAQTNTTNSG-DV 188
 QY 185 KALIFADAAARGSLNLS-DYLLDVEAGFEIWMOGGIGSFSFVS 228
 DB 189 KNFENILRNKGYNMAAGQYVLSYQFTPEPTGSGTLNVAWSWTASI 233

RESULT 15

094218 PRELIMINARY; PRT; 238 AA.
 ID 094218
 AC 094218;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE XYLOGLUCAN-SPECIFIC ENDO-BETA-1,4-GLUCANASE PRECURSOR.
 OS Aspergillus aculeatus.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5053;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-KSM 510;
 RX MEDLINE=99102417; PubMed=9884411;
 RA Pauly M., Andersen L.N., Kauppinen S., Kofoed L.V., York W.S.,
 RA Albersheim P., Darvill A.;
 RT "A xyloglucan-specific endo-beta-1,4-glucanase from Aspergillus
 RT aculeatus: expression cloning in yeast, purification and
 RT characterization of the recombinant enzyme."
 RL Glycobiology 9:93-100(1999).
 DR EMBL; AF043595; AAD02275.1;
 DR InterPro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR ProDom: PD004316; Glyco_hydro_12; 1.
 KW Signal.
 FT SIGNAL 1 14 POTENTIAL.
 FT CHAIN 15 238 XYLOGLUCAN-SPECIFIC ENDO-BETA-1,4-
 FT GLUCANASE.
 SQ SEQUENCE 238 AA; 25158 MW; FCCA6746D9AEC1B1 CRC64;

Query Match 14.1%; Score 175.5; DB 3; Length 238;
 Best Local Similarity 26.8%; Pred. No. 1.3e-05;
 Matches 61; Conservative 35; Mismatches 93; Indels 39; Gaps 10;

QY 17 GEYRVQTNENWSSA---OQCLTINTATG---AWTVSTANFSGGNG-----GAPATYPSI 64
 DB 33 GDTFLVNDLMGESAGTGSQCTGVDSISGDTIAHHTSM-SMSGSSSVKSYVNAALTF--- 88
 QY 65 YKCGHNGNCTTKNVGMPIQISQIGSAVTSMS-TVOVSSGAYDYADYDIWINSPTTGQPN 123
 DB 89 -----TPQLNCISSTIPTWKWYSGSSIVADYADTFLEAFASGSSK-- 131
 QY 124 GTIIMITLNSRGCVQPPGS--QTATGVTVAHGHTWVWQGOOTSMKIISYVLPQATISIN 181
 DB 132 -YEIMVWLALGAGAPISSTGSTIATPTIAGVWVKLYSGPNCDDTYVSFVADSTTESFSG 190
 QY 182 LDKAIFADAAARGSLNTSDYLDVEAGFEINOGGCG-LGSNSFSVSV 228
 DB 191 -DLNDEFTTYLVNDEGVSDELYLTLEAGTEPFGSNAKLTVSEYSISI 237

Search completed: August 29, 2002, 16:37:53
 Job time: 963 sec

This Page Blank (uspto)

Db	112	YGNCAPIRTT-LPMHSSIGSAPSVSRVRYGNNGVYNAAYDMLDPIRRTNG-VNRIEIMI	169
Qy	130	WLNRCGYVPPGSGQTATGCVIYAGHTNANWOGQOTSKIIISYVLTPGATSIINLDKATFA	189
Db	170	WFNVGVPQPIGSPVGT-AHGGKSMVETWGTSGNSDNVISF-LAPSAISSMSFDVKD-FV	226
Qy	190	DAARGSLNTSD-YLLDVEAGFEIWMOGGOGIGSFSFVSYTSG	231
Db	227	DQAVSHGLATPDWLTYSIQGFEPWEGGTGLAVNFSFSAVNAAG	269

```

RESULT      5
US-09-216-295-22
: Sequence 22, Application US/09216295
: Patent No. 6268328
:
GENERAL INFORMATION:
: APPLICANT: Mitchinson, Collin
: APPLICANT: Wendt, Dan J.
: TITLE OF INVENTION: No. 6268328e1 Variant EGIII-Like Cellulase Compositions
: FILE REFERENCE: GC555
: CURRENT APPLICATION NUMBER: US/09/216,295
: CURRENT FILING DATE: 1998-12-18
: NUMBER OF SEQ ID NOS: 41
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 22
LENGTH: 381
:
TYPE: PRT
:
ORGANISM: Streptomyces lividans CelB
US-09-216-295-22

```

Query Match	34.1%;	Score 424.5;	DB 4;	Length 381;
Best Local Similarity	42.3%;	Pred. NO. 7e-32;		
Matches	94;	Conservative	32;	Mismatches 89;
				Indels 7;
				Gaps 7;

[illegible]

```

RESULT 6
US-09-216-295-23
: Sequence 23, Application US/09216295
: Patent No. 6268328
: GENERAL INFORMATION:
: APPLICANT: Mitchinson, Collin
: APPLICANT: Wendt, Dan J.
: TITLE OF INVENTION: No. 6268328e1 Variant Ecdrr-Like Cellulase Compositions
: FILE REFERENCE: GC555
: CURRENT APPLICATION NUMBER: US/09/216,295
: CURRENT FILING DATE: 1998-12-18
: NUMBER OF SEQ ID NOS: 41
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 23
: LENGTH: 260
: TYPE: PRT
: ORGANISM: Rhodothermus marinus
: US-09-216-295-23

```

Query Match	32.8%	Score 407.5	DB 4	Length 260
Best Local Similarity	40.1%	Pred. No. 1,6e-30		
Matches	89	Conservative	30	Mismatches 84; Indels 19; Gaps 7

QY	14	VODGEYVOTNEMNSA	OOCLTINTATGAMV	STANFSGTGC	GADATPYSIKG	CHWG--	71
DB	50	VAGGRYRINNMWGA	ETAOICIEVGLTGG	FFTRADHDGNN--	VAAAPAI	FGCHMAPA	107
QY	72	---NCTRKNNM	PIQISGSAVMSW	STVSSGADYAVY	DMTNS--PPTT	GQNGTE	126
DB	108	RAIRRCAR--	AGAVRAHEL-----	DVTPITTRG	NNAAIYIM	SPYNSGNGSG	159
QY	127	IMIMNSRG	VOPEFSGQATG	VTAGHTMWN	WQOQTSK	IIISYLTPG	186
DB	160	IMIMNMNMG	GVMPGGSG	SRVAT--VELAG	ATEVWYA--	DMDMNYI	217
QY	187	IFADAARG	SLTSDYLLD	VEAGFEI	WOGGQIG	SNFSVS	228
DB	218	FLDDAVARG	YIRPEWYI	HAVERTG	FEIMEG	AGCLRTAD	259

```

RESULT 7
US-09-216-295-5
: Sequence 5, Application US/09216295
: Patent No. 6268328
: GENERAL INFORMATION:
: APPLICANT: Mitchinson, Collin
: APPLICANT: Wendt, Dan J.
: TITLE OF INVENTION: No. 6268328e1 Variant Egitr-Like Cellulase Compositions
: FILE REFERENCE: GC555
: CURRENT APPLICATION NUMBER: US/09/216,295
: CURRENT FILING DATE: 1998-12-18
: NUMBER OF SEQ ID NOS: 41
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 5
: LENGTH: 259
: TYPE: PRT
: ORGANISM: Aspergillus aculeatus
US-09-216-295-5

```

Query Match	20.9%;	Score 260.5;	DB 4;	Length 259;
Best Local Similarity	33.9%;	Pred. No. 8.6e-17;		
Matches	82;	Conservative	21;	Mismatches 96;
				Indels 43;
				Gaps 10;

```

0Y 17 GEPYRQTEEMN---SSAOCCITINTANGAMTV-STA-NESSGCTGAPATPSIKGCHWG 71
Db 31 GYITINNMLNCKDAGSGGCTTIVSASAGTSWSTKKNWSGEGENSVS-----YAN 81
0Y 72 NCTTANVMPIQISOIGSAVTSWSTTOVSSGAYDAVDIMWNSPTTGGPNCNGEIMWL 131
Db 82 SGLTFENKRLVQSISOI-PTTARMSYDNTGIRA-DVAADLEFAADINHWTSGDYELMWL 139
0Y 132 NSRGVQVEPESOTANGVTVACHTNNVMQGOOTSMKITSYVLTPCATSINLDKAIFADA 191
Db 140 ARYGVOPIGSOIAT-ATVDQOTELMWLGANGSOKTYSFVAAPPIYSFOG-DVNDPEKYL 197
0Y 192 AA-RGSLNTSDYLLD-----EAGFEIMGGOGGLGSNFSV 236
Db 198 TONHGFAPASQYLITLQGTPEPTGCPATLSVNMMSASVQAGREPMQNGAGLAVNFS 257
0Y 227 SV 228
Db 258 TV 259

```

RESULT 8
US-09-216-295-16
; Sequence 16, Application US/09216295
; Patent NO. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.

OY 183 DLKAFD--AARGSLNTSDYLDVAGEFIRMOGGLGNSRFSVS 228
Db 214 NIRD-FYNYLADSKOMLSKTKYVSVFEGTEGTCQINISMDYV 260

RESULT 11

US-08-032-848C-12
Sequence 12, Application US/08032848C
Patent No. 5475101
GENERAL INFORMATION:
APPLICANT: Ward, Michael
APPLICANT: Clarkson, Kathleen A.
APPLICANT: Weis, Geoffrey L.
APPLICANT: Larens, Edward
APPLICANT: Lorch, Jeffrey D.
TITLE OF INVENTION: Purification and Molecular Cloning of
TITLE OF INVENTION: EG III Cellulase
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 180 Kimball Way
CITY: South San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/032,848C
FILING DATE: MAR 17 1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Horn, Margaret A.
REGISTRATION NUMBER: 33,401
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415 742-7356
TELEFAX: 415 742-7217
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 233 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-032-848C-12

Query Match 17.2%; Score 214.5; DB 1; Length 233;
Best Local Similarity 30.3%; Pred. No. 1.5e-12;
Matches 67; Conservative 33; Mismatches 90; Indels 31; Gaps 11;

OY 16 DGEYVOTNEMNSA---OQCLTINTATGA---WTVSTANFSGGTGAPATYPSLYK 66
Db 14 NKKYLFNNWKGDEIKMOQTIYNSPISGMNHPSPSTHSYK-----AYPSLYS 65
OY 67 GCHNGNTTKNVGMPIDISGSAVTSMTQVSSGAYDAVDYDWTNSTPTTQOPNGT- 125
Db 66 GWHHTACYTESSGLPIDLSKSKTSITVYTIKATGYMAAYDTPHTTDDKANDSSPTD 125
OY 126 EIMTWLSRSGVOPFGSQTATGTVVAGHTNWNVOG-----OQTSWKIISVLTTPGATSI 179
Db 126 EIMTWLNDT-NAGPAGYIEI-VFLGSSNNVFEKGMINMADNGGGMVFSFVHSGTNSA 183
OY 180 SNLDLKAIFADAAARGLNTSD--YLLDVEAGEFIRMOGCGG 218
Db 184 S-LNIRH-FTDYLVOTKQMSDEKISSVERGEIF--GGDG 221

RESULT 12
US-09-216-295-20

Sequence 20, Application US/09216295
Patent No. 6268328
GENERAL INFORMATION:
APPLICANT: Mitchinson, Colin
APPLICANT: Wendt, Dan J.
TITLE OF INVENTION: No. 6268328e1 Variant BcIII-Like Cellulase Compositions
FILE REFERENCE: GC555
CURRENT APPLICATION NUMBER: US/09/216,295
CURRENT FILING DATE: 1998-12-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 20
LENGTH: 246
TYPE: PRT
ORGANISM: Emericella desertoru
US-09-216-295-20

Query Match 16.8%; Score 208.5; DB 4; Length 246;
Best Local Similarity 31.3%; Pred. No. 5.8e-12;
Matches 71; Conservative 41; Mismatches 80; Indels 35; Gaps 12;

OY 17 GEYVOTNEM---NSSAOCLTINTATG---AATVSTANFSGGTGAPATYPSLYKGM 70
Db 41 GNFIVYNNLMGQDNADSGOTGVDSANGNSISMH-TTWSMGSGSSVYS-----YANAAY 94
OY 71 GNCCTKNVGMPIQISGSAVTSN---STQVSSGAYDAVDYDWTNSTPTTQOPNGTE 126
Db 95 QFTSTK-----LNSLSITPSMKWQSTTDIYA---NVAIDFTSS---SAGDSLEY 141
OY 127 IMIWLNSRGVQPF---GSQTATGTVVAGHTNWNVOGQTSWKIISVLTTPGATSI 183
Db 142 IMIWLALGAGAPISSTGSIAT-VTLGQVWMSLYSGPNSGMQYSEFVASTTEESF-AD 199
OY 184 LKAIFADAAARGLNTSDYLDVAGEFIRMOGCGG-LGSNFSVSYSY 229
Db 200 LMDFINVLAENQGLSSQYLTHVQAGTEPFTGTATLTVSSYSVS 246

RESULT 13
US-09-216-295-12
Sequence 12, Application US/09216295
Patent No. 6268328
GENERAL INFORMATION:
APPLICANT: Mitchinson, Colin
APPLICANT: Wendt, Dan J.
TITLE OF INVENTION: No. 6268328e1 Variant BcIII-Like Cellulase Compositions
FILE REFERENCE: GC555
CURRENT APPLICATION NUMBER: US/09/216,295
CURRENT FILING DATE: 1998-12-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 12
LENGTH: 238
TYPE: PRT
ORGANISM: Fusarium equiseti
US-09-216-295-12

Query Match 16.6%; Score 206.5; DB 4; Length 238;
Best Local Similarity 31.7%; Pred. No. 8.5e-12;
Matches 71; Conservative 28; Mismatches 86; Indels 39; Gaps 14;

OY 8 ONGVTSVDDGEYVOTNEM---NSSAOCLTINTATGA-WTVSTANFSGGTGAPATY 61
Db 23 QYGLIS-SDG-YSLNNVWKGDSGTGDCCTHVNNMNAAGMDVEW-NMSGKDNV-KSY 78
OY 62 PSIRKGCHMGCTKNVGMPIQISGSAVTSMTQVSSGAYDAVDYDWTNSTPTTQOPNGTE 113
Db 79 PN-----SALLIGEDKKTISITNMQSTAKMKSGDWLRADVADYDFT 122
OY 114 STPTTQOPNGTEIMIWLNSRGVOPFGSQTATGTVVAGHTNWNVOGQTSWKIISVLT 173

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:16:25 ; Search time 127 Seconds

(without alignments)
1074.005 Million cells updates/sec

Title: US-09-917-384-1

Perfect score: 6525

Sequence: 1 MERTQSGRCRCRYRGTTTMM.....GFGNSYSGTNAAPLTCTAS 1228

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq.032802:*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1023	15.7	1010	19	AAW34989 Tereidinbacter end
2	1000	15.3	596	17	AAW34989 Tereidinbacter end
3	999	15.3	423	17	AAW34989 Tereidinbacter end
4	708.5	10.9	1751	20	AAW13493 Truncated cellulase
5	668.5	10.2	381	21	AAW06368 Streptomycetes livid
6	663.5	10.2	1426	21	AAW13492 Rhodothermus marin
7	651.5	10.0	429	21	AAW13492 Truncated cellulase
8	622.5	9.5	386	21	AAW67497 Amino acid sequenc
9	620	9.5	371	21	AAW06367 Streptomycetes sp. E
10	620	9.5	371	21	AAW06367 Streptomycetes sp. E
11	620	9.5	371	21	AAW14880 Streptomycetes livid

12	620	9.5	371	21	AAW84345
13	618	9.5	371	21	AAW08473
14	618	9.5	371	21	AAW67496
15	574.5	8.8	395	20	AAW84797
16	573	8.8	473	21	AAW01077
17	565.5	8.7	476	19	AAW44853
18	565.5	8.7	476	19	AAW44853
19	548	8.4	471	6	AAW50308
20	547	8.4	521	22	AAW84788
21	543	8.3	521	17	AAW89927
22	543	8.3	521	22	AAW84786
23	543	8.3	521	22	AAW84787
24	543	8.3	562	21	AAW69508
25	540	8.3	471	16	AAW7262
26	540	8.3	471	16	AAW02025
27	535	8.2	457	18	AAW25789
28	493.5	7.6	5179	22	AAW24516
29	488	7.5	1853	19	AAW43108
30	485.5	7.4	446	20	AAW01076
31	456	7.0	531	18	AAW15238
32	444.5	6.8	449	19	AAW56739
33	431.5	6.6	432	21	AAW18414
34	427.5	6.6	782	12	AAW15625
35	419	6.4	260	20	AAW06369
36	419	6.4	260	21	AAW14882
37	419	6.4	260	21	AAW84347
38	417.5	6.4	551	18	AAW18790
39	413.5	6.3	616	20	AAW13494
40	412	6.3	459	19	AAW56738
41	409	6.3	493	20	AAW28850
42	409	6.3	493	20	AAW43218
43	404	6.2	476	21	AAW54123
44	399.5	6.1	406	22	AAW37250
45	386	5.9	1352	22	AAW63962

ALIGNMENTS

RESULT 1	AAW34989 standard; Protein: 1010 AA.
XX	AAW34989;
XX	21-MAY-1998 (first entry)
DE	Tereidinbacter endoglucanase.
XX	
XX	Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
KW	biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW	thermostable enzyme; thermophilic; glycosidase.
XX	
OS	Tereidinbacter sp. (Clone 42GP1).
XX	
PN	W09744361-A1.
XX	
PD	27-NOV-1997.
XX	
PE	22-MAY-1997; 97WO-US08793.
XX	
PR	22-MAY-1996; 96US-0651572.
XX	
PA	(RECO-) RECOMBINANT BIOCATALYSIS INC.
XX	
PI	Lam DE, Mathur EJ;
XX	
DR	WPI; 1998-018435/02.
DR	N-PSDB; AAT94197.
XX	
PT	Endoglucanase(s), preferably form archael bacterium, AEPIT 1a -
PT	useful to degrade carboxymethylcellulose and hydrolyse of
PT	beta-1,4-glycosidic bonds in cellulose

Amino acid sequenc
Actinomycetes sp. 3
Actinomycetes cell
Amino acid sequenc
H. insolens cell6A
Humicola insolens
H. insolens cellu
Cellulobiohydrolase
Acidothermus cellu
A. cellulolyticus
Acidothermus cellu
Acidothermus cellu
Acidothermus cellu
T. longibrachiatum
Trichoderma cellob
Acromonium cellulo
C899P predicted am
C. thermocellum ce
H. insolens Cell6B
Scaffoldin protein
Orpinomyces cellu
Amino acid sequenc
Cellulase AE-1. A
Rhodothermus marin
Emeritella deserto
Amino acid sequenc
Corrected Bacillus
Truncated cellulase
Orpinomyces cellu
Pectate lyase-link
A mannase-linker
Novel human diagno
Amino acid sequenc

XX Claim 1; Fig 1E; 164pp; English.

PS This protein comprises an endoglucanase of *Teredinibacter* (clone

CC 42CP1) that is capable of degrading carboxymethylcellulose and of

CC hydrolysing the beta-1,4-glycosidic bonds in cellulose. It has

CC homology to an endoglucanase of archaeobacterium AHP11a (see

CC AAM34985). It can be produced from native cells or from recombinant

CC host cells, especially prokaryotic host cells transformed with a

CC plasmid or virus-derived vector including the endoglucanase DNA

CC (see AAT94197). 24 Endoglucanases (see AAM34986-W35008) are claimed.

CC They can be used to degrade cellulose for the conversion of plant

CC biomass into fuels and chemicals, for use in detergents, textiles,

CC animal feed, waste treatment, and in the fruit juice and brewing

CC industries for the clarification and extraction of juices.

XX Sequence 1010 AA;

SO

Query Match 15.7%; Score 1023; DB 19; Length 1010;

Best Local Similarity 48.5%; Pred. No. 6.5e-41;

Matches 210; Conservative 68; Mismatches 121; Indels 34; Gaps 13;

QY 57 VDNPPAGA-TFEVNPYMAOEVSANQTNATLAKMRVSTYSTAVWMDRIATNGVNG 115

DB 598 vdnppagaqkyinpmws-----asaanepggs-----vianepsfvmdrigalegpad 647

QY 116 GPGITTYIDALISQOOGTTPREYIEIYIDLPGRCALALASNGELPATAGLQTYETQYTD 175

DB 648 gmgldrhnealaq----gadlfmfvvyd]purdcaalasangelsedgfnlydsdyla 703

QY 176 PLASLSPKXSLRIYVILIEPDSLPMATNMSIQAC--ATNVPYVEGIEFALTKLHAI 233

DB 704 plvelisdpayagiklaavlevisipnvtlnsepdceangpgyrdgdirhatelgk1 763

QY 234 PNVTYINDAASHGWLGMPPNN-ASGYVOEVQKYLNASIGVNGIDGVVTANTPTLKEPPM 292

DB 764 pnvytyvdalshgwlgmndfnagynlllyevvaangsginplagtsnasytprpcep1 823

QY 293 -TATQOVGGQPVESANFYQWNPDIADYADVLYSRLVAAGPSSIGMLIDTLRNGWGP 351

DB 824 pdanfygvgqpyrsdifyewasylakpfvldwrsamiskgmspsigmlidtlarnwg9p 883

QY 352 NEPTGPSTATDVTNFTVNSKIDLRHGMCMQNGAGLGQPPQASPTDEPNHLDAYWI 411

DB 884 erptagsstnnlntfvesrldtrhrgmwcnpq-gvgyrptlaa---pspgidayvvw 938

QY 412 KPPGESDGTSAAS--DPT-TGKKSDDPMCDP---TYTTSYGLVNLALPNSPIAGQMPFA 463

DB 939 kpggesdgvsdpnfeldpnndpukqhdpmcdpfasnsnsayg--tganpnphagrwfpe 996

QY 464 QPDQLVANNRPAY 476

DB 997 afglllenappt 1009

RESULT 2

AAR90715

ID AAR90715 standard; Protein; 596 AA.

XX AAR90715;

AC

XX 06-APR-1996 (first entry)

DT

XX

DE Thermostable cellulase-E3.

XX

KW Cellulase-E3; Thermomonospora fusca; thermostable; plasmid ps24;

KW plasmid ps26; *Escherichia coli*; cloning; *Streptomyces lividans*;

KW papain; catalytic domain; cellulose; chitosan; protease resistance;

KW synergism; cellolobiondrolase; beta-glucosidase; saccharification;

KW surfactant additive; paper recycling; delinking; paper refining.

OS Thermomonospora fusca strain YX36.

XX Key Location/Qualifiers

EH Peptide 1..38

ET Peptide /note= "signal peptide"

ET Peptide 39..48

ET Domain /note= "N-terminal peptide (AAR90717)"

ET Domain 174..596

XX /note= "Catalytically active domain (AAR90716)"

PN W09600281-A1.

PD 04-JAN-1996.

PP

XX 23-JUN-1995; 95MO-US09069.

XX

PR 24-JUN-1994; 94US-0265429.

XX

PA (CORR) CORNELL RES FOUND INC.

XX

PI Walker LP, Wilson DB, Zhang S;

XX

DR WPI; 1996-068865/07.

DR N-PSDB; AAT15595.

XX

PT Cellulase from *Thermomonospora fusca* - is thermostable and is useful

PT in a variety of industrial applications e.g. clarification of fruit

PT juices, fabric softening, etc

XX

XX Claim 1; Page 32-37; 53pp; English.

CC The sequence represents thermostable cellulase-E3 (EC-3.2.1.4) from

CC *Thermomonospora fusca*, and is encoded by inserts in plasmid ps24

CC and plasmid ps26, obtained by cloning in *Escherichia coli*. The

CC gene may be cloned and expressed in *E. coli* or *Streptomyces*

CC *lividans* for recombinant cellulase-E3 production. Cellulase-E3,

CC or a catalytically active polypeptide obtained by papain digestion,

CC hydrolyses cellulose or chitosan at pH 5-11 and 40-70 deg C, and

CC has significant activity at over 60 deg C. E3 has higher stability

CC to proteolysis in culture supernatant than *T. fusca* cellulases-E2

CC and -E5, and shows strong synergistic activity when combined with

CC other cellulases, cellolobiondrolase and optionally beta-glucosidase.

CC The cellulase and mixtures may be used e.g. in cellulose

CC saccharification for ethanol production, fruit juice clarification,

CC as a surfactant additive, or in delinking or refining of recycled

CC paper.

XX

SO Sequence 596 AA;

QY 50 PAIAATHVDNPAGATFEVNPYMAOEVSANQTNATLAKMRVSTYSTAVWMDRIAT 109

DB 172 ptnpgekvampiegaklynpwpsakaapegs-----avaestaavwldrlga 221

QY 110 ING---VNGPGGLTTYIDALISQOOGTTPREYIEIYIDLPGRCALALASNGELPATAG 165

DB 222 iegndsptgsmglrdhleeav-rqsggqplrlqyvllylnlprdcalaasangelpde- 278

QY 166 LOTYETQYIDPLASLISN-PKTSRLRIYVILIEPDSLPMATNMSIQACATAPVY----- 218

DB 279 ldrkyseyldpladlmwdfadyenlrivaltidsipnlvtvnvgngyrgelcaymknqg 338

QY 219 YEOGIEYALTKLHAIPNVIYMDAASHGWLGMPPNNASGYVOEVQKYLNAS-IGVNGIDOF 277

DB 339 yngvgyalrklqelpnyvnyidaahngwlgwdsnngpsvdlfiyeaanagsstvdgvygf 398

QY 278 VNTNATVYTLKPPFMATQOVGGQPVESANFYQWNPDIADYADVLYSRLVAAGPSSIG 337

DB 399 lantanysatvepyldvngvtngqllrsgskwvdmngyvdelsfvqdlrallakfrsdi 458

Query Match 15.3%; Score 1000; DB 17; Length 596;

Best Local Similarity 47.7%; Pred. No. 4.6e-40;

Matches 210; Conservative 55; Mismatches 143; Indels 32; Gaps 9;

QY 338 GMLDITLRNGCGPNEPPTSTATDVNTFVNOSKIDLRQHRGLMCNONGAGLGQPPQASP 397
 DB 459 gmlldtsrnyg9pnrlptgssadlntlydesrltrihpwnwcnqagaglgcrptvnp 518
 QY 398 TDFPNALHDAYWIKPPGSDGTSAASDPTTGKKSDPMCDPRTYTSYGLTN---ALPNS 454
 DB 519 a---pyvdahyvwvkppgsdgaeeipndegkyfdlrmcdpkygnarnnpsgalpna 574
 QY 455 PIAGOMFPAQFDLVANARP 474
 DB 575 plsgnhfsaqfrelanayp 594

RESULT 3
 AAR90716
 ID AAR90716 standard; Protein; 423 AA.
 AC AAR90716;
 DT 06-APR-1996 (first entry)
 DE Thermostable cellulase-E3 catalytic domain.
 KM Cellulase-E3; Thermomonospora fusca; thermostable; papain;
 KM catalytic domain; Escherichia coli; cloning; Streptomyces lividans;
 KM cellulose; chitosan; protease resistance; synergism;
 KM cellobiohydrolase; beta-glucosidase; saccharification;
 KM surfactant additive; paper recycling; delinking; paper refining.
 OS Thermomonospora fusca strain YX36.
 PN WO9600281-A1.
 PD 04-JAN-1996.
 PF 23-JUN-1995; 95WO-US09069.
 PR 24-JUN-1994; 94US-0265429.
 PA (CORR) CORNELL RES FOUND INC.
 PI Walker LP, Wilson DB, Zhang S;
 DR WPI; 1996-068865/07.
 N-PSDB; AAT15596.
 PT Cellulase from Thermomonospora fusca - is thermostable and is useful
 in a variety of industrial applications e.g. clarification of fruit
 juices, fabric softening, etc
 PS Claim 1; Page 38-40; 53pp; English.
 XX The sequence corresponds to a catalytically active domain of
 CC thermostable cellulase-E3 (EC-3.2.1.4) from Thermomonospora
 CC fusca, and is isolated by papain cleavage of the full-length
 CC protein (AAR90715). The binding domain and linker region have
 CC been removed from the full-length protein. The protein may be
 CC produced in recombinant form in Escherichia coli or Streptomyces
 CC lividans. The domain retains full activity, but shows reduced
 CC cellulose binding. The catalytic domain, or the corresponding
 CC full-length cellulase, hydrolyses cellulose or chitosan at pH
 CC 5-11 and 40-70 deg C, and has significant activity at over 60
 CC deg C. E3 has higher stability to proteolysis in culture
 CC supernatant than T. fusca cellulases-E2 and -E5, and shows strong
 CC synergistic activity when combined with other cellulases,
 CC cellobiohydrolase and optionally beta-glucosidase. The cellulase
 CC and mixtures may be used e.g. in cellulose saccharification for
 CC ethanol production, fruit juice clarification, as a surfactant
 CC additive, or in delinking or refining of recycled paper.
 SO Sequence 423 AA;

Query Match 15.3%; Score 999; DB 17; Length 423;
 Best Local Similarity 48.3%; Pred. No. 3,5e-40;
 Matches 209; Conservative 55; Mismatches 137; Indels 32; Gaps 9;

QY 57 VDNPRYGAATFFVNPYNAQEQSEANQNTNTLAAKHNVSTYSTAYWMDIAANG---- 112
 DB 6 vdnprfygaktynpvnvsakaapegys-----avaneslavldtligalegndsp 55
 QY 113 VNGCPGTTTYLDAALSOQOQTPEVEIEIYIDLPGRDCALALASNGELPATNAGLOTETQ 172
 DB 56 ttsgmglrdhleeav-rqsgqdpiltqivlylnlpgtdcaalaangelgpd---ldrykse 112
 QY 173 YIDPIASILSN-PKYSRLRVITIEPDSLPAVNTNMSIQACATAPY-----YEOGIEY 225
 DB 113 yidpiadimwdfadyenlrlvaltelalsplnlvtngvgngtelcaymknggyvngvy 172
 QY 226 ALTKLAIPRVITYMDAAHSGMIGWRPNASGVQEQKVLNAS-IGVNGIDGFTVNTANT 264
 DB 173 alrklgeiprvnyidaahgwlgswnlfgpsvdlfyeeanaagstvdyvhgflsnlany 232
 QY 285 TPKEPFMTATQOVGQGVESANFYQWNPDIADADVAVDLYSRLVAAFPSSIGMLITDL 344
 DB 233 salvepyldvngtlvngqlirgskwvdwnqyvdelstfvqdlrgaliaqfisdgmlldts 292
 QY 345 RNCWGPNEPFGSTATDVNTFVNOSKIDLRQHRGLMCNONGAGLGQPPQASPTDFPNAH 404
 DB 293 rncwgpnpurptgssadlntlydesrltrihpwnwcnqagaglgcrptvnpa---pg 348
 QY 405 LDAYWVWIKPPGESDGTSAASDPTTGKKSDPMCDPRTYTSYGLTN---ALPNSPIAGOMF 461
 DB 349 vdayvwvkppgsdgaeeipndegkyfdlrmcdpkygnarnnpsgalpnaplsgnhwf 408
 QY 462 PAQFDLVANARP 474
 DB 409 seqfrelanayp 421

RESULT 4
 AAY13493
 ID AAY13493 standard; Protein; 1751 AA.
 AC AAY13493;
 DT 30-JUL-1999 (first entry)
 DE Truncated cellulases comprising amino acid sequence.
 KM Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
 KM Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
 KM cotton-containing fabric; stonewashing.
 OS Unidentified.
 PN EP921188-A2.
 PD 09-JUN-1999.
 PF 15-SEP-1998; 98EP-0810919.
 PR 19-SEP-1997; 97US-0932571.
 PA (CLRN) CLARIANT FINANCE BVI LTD.
 PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;
 DR Gibbs MD, Morgan H, Williams DP;
 N-PSDB; AAX55662.
 PT New truncated cellulase proteins, useful in detergents and for
 producing 'stonewashed' denim
 PS Claim 7; Page 37-41; 65pp; English.

XX The invention relates to a recombinant cellulase active protein free of
 CC proteinases of native thermophilic and alkaliphilic origin, comprising
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel E1/2/3,
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
 CC length sequences, or functional equivalents. Cel B5 extends from amino
 CC acid A1011 to P1424 or N1425 or N1426, and Cel B4/5 extends from amino
 CC acid K635 to N1426 in the sequence shown in AY13492. Cel E1 extends
 CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
 CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
 CC K1751 and the stability region extends from amino acid E482 to G635 in
 CC the sequence shown in AY13493; Cel E3/B5 is shown in AY13494. The new
 CC enzymes are useful in laundry detergent compositions to prevent or
 CC remove staining, backstaining or graying, for use on cellulosic
 CC materials including cotton-containing fabrics. They are especially useful
 CC for preventing redeposition of colorant during stonewashing, and for
 CC processing of textiles where cellulose breakdown is required. The new
 CC truncated enzymes show reduced redeposition of dye compared to using
 CC non-truncated cellulase compositions.

XX Sequence 1751 AA:

Query Match 10.9%; Score 708.5; DB 20; Length 1751;
 Best Local Similarity 22.3%; Pred. No. 1,1e-25;
 Matches 304; Conservative 165; Mismatches 429; Indels 467; Gaps 54;

QY 35 AVSTAASTVPLAMOHRAIAATHV-----DNPYAGATFEVN----- 69
 DB 192 AASIAASVAVLKERNQAAYLQHAKLLEFADTTRSDAGYTAATGYTSGFIDDLGW 251
 QY 70 -PYV-----AOEYQSEANOTNA-----TLAKMNVSTYSTAV 102
 DB 252 AAVVLYLATNDSSYLTKAEELMSEYANGTCWCDDVRYGLIMLAKIGKELYKAV 311
 QY 103 -----WMDRIAANGVNGSGPLTYYLDALSOOGTPEYIEIYTYDLPGRDC----- 150
 DB 312 ERNDHWIRLTLC-----PKGMAYLTQWGLSYATTAELACYADWSGCDENKKTKY 365
 QY 151 -----AALASNGELPATAAGIQYETQYIDPFIASILSNPKYSSRIYVITIEPDSL 201
 DB 366 LNFKSGQIDYALGSLTGRSFVVG-----LNY-----PQH-----P 396
 QY 202 NAYTNMSIOACATAVPYEYEGIEYALTKLHAIPNYIYNDAHSHWGLMPNNASGYQVEV 261
 DB 397 HNRHNSWASMKIPEYHNLHY-----GALVGGPSDDSDYNDDI 437
 QY 262 QKVLNASIGVNGIDGFVNTANTYPLKEPFMTATCOVGOPVESANFYQWNPDIIDEADYA 321
 DB 438 TDYVQNEVACDYNAGIVGALAK-----MYGLYGGPDIIDFKALE-TPNDDEL-FV 485
 QY 322 VDLYSRLVAAQFPSSSIGMLIDTLRNGWGRNEPPEPSTATDVNTFVNSKIDLRQHRGLW 381
 DB 486 ESKIGNAGGPRYTEVLSYLYN-RTGW-----PRVTKLSEF-KYFIDLEI----- 529
 QY 382 CNONGAGIGOPPOASPTDFPNAHLDA-----YVW-----IKRPGESD 418
 DB 530 -----IQAGYSPDVVKVQ--TYVLEAGKISGPLYWDKRNILYVLDVFSGTKIIPPGVE 582
 QY 419 GTSASAPDTTKKSDPMCDPTTYSYGLVINALPNSPIAGOMFPAQFD--OLVANARAV 476
 DB 583 HKKGAGFKLAVPGGYPW-dpLNDPEYKGLTQLEKNK-----YIAAYHNNHLVWGLPGE 636
 QY 477 PTSTSSSPPPPPPSASPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPS 536
 DB 637 ATST-----PAT 644
 QY 537 SSPSPSPSPSSPSPPSSPSPPSPSPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSS 595
 DB 645 STPT 686
 QY 596 DSAPGDNDQIKPGLQLVNNTGSSSVDLSTVRYVWFTRDGSGTLVYVNCWAMGCGNITAS 655

DB 687 ETASSTGSLRPFKLVNGSSSVDSLSTVRYVWFTVGDGKQSAV-cWAGIAGSNVTFN 745
 QY 656 FGSVNPATPATDYVQLQSLFTGCT--LAAGSTGEIQNRVNSDMNSFETNDYSY-GTNT 712
 DB 746 FVKLSYSGSVGGLYEVGFSSAGQLQPGKDTGDIQVRFNKNDSYNGQADDWLQGMT 805
 QY 713 TFGQWTKYTVVNVNVLWGTGTEPSSGTSPP-----TPSPSPSPSPSGDVTTP 760
 DB 806 NYGENAKVLYVDGLVWVGEGGALPAPTATPCTPCTPCTPCTPCTPCTPCTPCTPCT 865
 QY 761 SVPTGLVVTGVSGLSSVSLAMNASTDNGVAVHNYRNGLVGLGPTVNSFTDTGLAAGTAY 820
 DB 866 PAPCT--ASPYGGS-----YCTPSE----- 882
 QY 821 TYTYAAV-DAAGNTSAPSTPYATATTSPSPSPPTGTTCCTCPGPNQNGVTSVQGDYR 879
 DB 883 SYGALKVYVANGUNLSPCT-----HVLNPKIKLENVGTLAVLS-----R 921
 QY 880 VQTNW---NSSAOCLINTATGA--WTYSTANFSSGTGAPATYPSYK---GCHMGN 931
 DB 922 VKVRYWYCLDGAATQSVSVASSINPAYIDVKFKLGANAGADYVVELGFKSGAVLAAG 981
 QY 932 CTTKNVGMPIQISQISAVTSMSTTOVSSGAYDAVDIMTNSPTP-----TTG----- 979
 DB 982 GSTKEIRLSI-----QKSGSGYNGNDYSVTSANSYLENEKVTGYIDDLV 1026
 QY 980 ---OP-NGTEIMTW-----LNSRQVQPFSGGTATGTVYAGHTWVWVQGOOT 1022
 DB 1027 VWGREPGFNAQIKVYVANGUNLGSMTVNLNPKIKLENVGTLAVLSRVKRYWYLTGAT 1086
 QY 1023 SMKTIISYVLTFGATSIWLDKAIFFADAAAGSLNTSPYLLDY--ENGFELWQGO----- 1076
 DB 1087 QSVSVLTSSINP-----AYIDVKFKVLGANAG--ADYVVELGFKSGAVLAAGQSTKE 1137
 QY 1077 -----GLG-----SNSFSVSVTSG-----TSPTPSP 1098
 DB 1138 LRLSIQKSGSGYNGNDYSVTSATGYLENEKVTGYIDGALVWGREPSTGKTPAGVVPCTP 1197
 QY 1099 SPTPTPSPPTPTPSPSPPTPSPSPSSSSGVACRATVYVNSDMSGFTATVYVNTTGSRA 1158
 DB 1198 APT 1242
 QY 1159 TNGWTVAMSPFGNQTIVTNYWMTALQSGASVATATNLSTNNYIQRP 1203
 DB 1243 RTKIS-PLYIANGQDIGVNHPRRLTGNRLTYGWE-nmmsnag 1285

RESULT 5
 AAY06368
 ID AAY06368 standard; Protein: 381 AA.
 XX
 AC AAY06368;
 XX
 DT 06-SEP-1999 (first entry)
 XX
 DE streptomyces lividans EGIII-like cellulase.
 KW Cellulase; endoglucanase; EGIII; textile; feed additive; baking;
 KW food processing; grain wet milling; pulp; paper.
 XX
 OS Streptomyces lividans.
 XX
 FN W09931255-A2.
 XX
 PD 24-JUN-1999.
 XX
 PF 14-DEC-1998; 98WO-US26552.
 XX
 PR 16-DEC-1997; 97US-0991720.
 XX
 PA (GENAV) GENENCOR INT INC.
 XX
 PI Bower BS, Fowler T, Phillips JT;

XX DR WPI; 1999-395187/33.
XX PT EGIII like cellulase
XX PS Example: Fig 6; 47pp; English.
XX
CC The present polypeptide represents a full-length sequence of a
CC novel EGIII-like cellulase of *Streptomyces lividans* (Ce1b). It was
CC deduced from a gene sequence isolated from genomic DNA using PCR
CC primers (see AAX59180-91) based on conserved motifs (see AAY06325-29)
CC of *Trichoderma reesei* EGIII cellulase and related enzymes. PCR
CC has been used to identify novel EGIII-like enzymes, including the
CC present protein, from bacterial and fungal sources (see AAY06331-70).
CC The sequence shows homology to T. reesei EGIII (see AAY06330). Also
CC provided by the invention are vectors, host cells and methods
CC for the recombinant production of such enzymes, which can be used
CC in the treatment of cellulose-containing textiles, as feed
CC additives, in the treatment of wood pulp, in the reduction of
CC biomass to glucose, in the stone washing of indigo dyed denim, or
CC as laundry detergent components (all claimed).
XX
SQ Sequence 381 AA:

Query Match 10.2%; Score 668.5; DB 20; Length 381;
Best Local Similarity 39.1%; Pred. No. 1.7e-24;
Matches 158; Conservative 49; Mismatches 150; Indels 47; Gaps 13;

OY 829 MAGNTSAPSPRYATTTSPSPPTPTGTYTDCPTPCPNQGVTSVQCDERYVQTNEMNSS 888
DB 17 algaavlafalvaslvtlaapagadt---flicepr---fgtltlqg-ryvvqnnrvgst 67
OY 889 AAOCLTINTATGAMTVSTANFSGTGAPATYPSIRKCHMGNCTRNKVMPIQISIGS 948
DB 68 apqcvta-ldtg-frrvqdgdsaprtngapksypsvfngchytnc-spgtdlprldtvsaa 124
OY 949 AVTSMSTTQVSSGAYDAVDIWTNSTPTTGGPNGTETIMIMLNSRGCVOPFGSQTATGVT 1008
DB 125 apaslaygfvdgavynasylwdlptarldg-vnqteimlwfvrvgplqigsrvgt-as 182
OY 1009 VAGHTMNVWMOGQOTSMKIIISYVLPGATISINLDKAIKIPDAARGLSNTSDVLDYEAG 1068
DB 183 vvgitvevwsygnsgndvlsfv-apsalsgwsldvmdfvratvarglaendwyltsvqag 241
OY 1069 FEIMOGGOGIGSNSFSVSVSTSGTSSPTSPSPPTPTSPSPPTSPSPSPSSSG 1128
DB 242 fepwqngaglaavnafstsvetgtptgt-----dpgdpggpps--- 277
OY 1129 VACRATVYVNSDMGSGFTATVTNTGSRATNGMTVAMSGNQTVYNNVTALTOGSGAS 1188
DB 278 -acavsygtlnv-wqdgftadvtvntgtlapvdgwfqlfplpsqgrltnaanaaltpssgs 335
OY 1189 VTATNLSYNNVIOPGOSTTGFNGSGTNAAPT-----LSCST 1226
DB 336 vlatgashnarlapggsifgfglyggafaeptgfrlntact 379

RESULT 6
AAB14881
ID AAB14881 standard; Protein; 381 AA.
XX
AC AAB14881;
XX
DT 21-NOV-2000 (first entry)
XX
DE Rhodothermus marinus EGIII-like cellulase.
XX
XX Rhodothermus marinus; EGIII-like cellulase.
XX
XX Rhodothermus marinus; Trichoderma reesei; endoglucanase III; EGIII;
XX
XX cellulase; mutant; enzyme stability; textile treatment;
XX
XX wood pulp treatment; feed additive; detergent.
OS Rhodothermus marinus.

XX PN WO200037614-A2.
XX PD 29-JUN-2000.
XX
XX PF 12-NOV-1999; 99MO-US26704.
XX
XX PR 18-DEC-1998; 98US-0216295.
XX
XX PA (GEMV) GENENCOR INT INC.
XX
XX PI Mitchinson C, Wendt DJ;
XX
XX WPI; 2000-482483/42.
XX
CC Novel endoglucanase III or endoglucanase III-like cellulase useful for
CC treating textiles and wood pulp comprises a substitution or deletion at
CC specified positions in the wild form of endoglucanase III -
XX
XX Example 1; Fig 3; 52pp; English.
XX
CC The present sequence is a cellulase related to endoglucanase III (EGIII)
CC from *Trichoderma reesei*. EGIII-like genes were isolated from genomic DNA
CC libraries constructed from various microorganisms by PCR. The isolated
CC genes showed significant homology to EGIII from T. reesei. Certain
CC substitution and deletion mutations have been incorporated into EGIII and
CC EGIII-like cellulases to produce variant enzymes with improved stability,
CC e.g. increased resistance to temperature stress. The mutants may be used
CC in textile and wood pulp treatment, as a feed additive, and for reducing
CC biomass to glucose. They are also useful for stonewashing or indigo dyed
CC denim and as an agent in laundry and dish detergents.
XX
SQ Sequence 381 AA:

Query Match 10.2%; Score 668.5; DB 21; Length 381;
Best Local Similarity 39.1%; Pred. No. 1.7e-24;
Matches 158; Conservative 49; Mismatches 150; Indels 47; Gaps 13;

OY 829 MAGNTSAPSPRYATTTSPSPPTPTGTYTDCPTPCPNQGVTSVQCDERYVQTNEMNSS 888
DB 17 algaavlafalvaslvtlaapagadt---flicepr---fgtltlqg-ryvvqnnrvgst 67
OY 889 AAOCLTINTATGAMTVSTANFSGTGAPATYPSIRKCHMGNCTRNKVMPIQISIGS 948
DB 68 apqcvta-ldtg-frrvqdgdsaprtngapksypsvfngchytnc-spgtdlprldtvsaa 124
OY 949 AVTSMSTTQVSSGAYDAVDIWTNSTPTTGGPNGTETIMIMLNSRGCVOPFGSQTATGVT 1008
DB 125 apaslaygfvdgavynasylwdlptarldg-vnqteimlwfvrvgplqigsrvgt-as 182
OY 1009 VAGHTMNVWMOGQOTSMKIIISYVLPGATISINLDKAIKIPDAARGLSNTSDVLDYEAG 1068
DB 183 vvgitvevwsygnsgndvlsfv-apsalsgwsldvmdfvratvarglaendwyltsvqag 241
OY 1069 FEIMOGGOGIGSNSFSVSVSTSGTSSPTSPSPPTPTSPSPPTSPSPSPSSSG 1128
DB 242 fepwqngaglaavnafstsvetgtptgt-----dpgdpggpps--- 277
OY 1129 VACRATVYVNSDMGSGFTATVTNTGSRATNGMTVAMSGNQTVYNNVTALTOGSGAS 1188
DB 278 -acavsygtlnv-wqdgftadvtvntgtlapvdgwfqlfplpsqgrltnaanaaltpssgs 335
OY 1189 VTATNLSYNNVIOPGOSTTGFNGSGTNAAPT-----LSCST 1226
DB 336 vlatgashnarlapggsifgfglyggafaeptgfrlntact 379

RESULT 7
AAV13492
ID AAV13492 standard; Protein; 1426 AA.
XX
AC AAV13492;

XX Novel variant endoglucanase III-like cellulases with improved
 PT surfactant stability and resistance to temperature stress, useful for
 PT textile processing or cleaning, treating wood pulp, food and grain, and
 PT reducing biomass to glucose
 XX
 PS Disclosure: Page 65-66; 73pp; English.
 XX
 CC The present sequence represents an endoglucanase III (EGIII)-like
 CC cellulase. The cellulase has homology to the Trichoderma reesei EGIII
 CC protein. The variant cellulases have improved temperature stability,
 CC and improved surfactant stability. The variant cellulases and
 CC compositions containing them are used in textile processing or cleaning,
 CC e.g. stonewashing of indigo dyed denim, and modifying the texture, feel
 CC or appearance of cellulose containing fabrics (e.g. improving fabric
 CC smoothness or removing pills and fibrils). The compositions may also be
 CC used for the removal of immature or dead cotton from cellulosic fibres
 CC or fabric, which can cause uneven dyeing. The cellulase may also be used
 CC in a detergent composition for washing laundry and dishes and in the
 CC treatment of animal feed, wood pulp, paper, non-animal foods and grains.
 CC The enzymes may also be used in the reduction of biomass to glucose.
 CC
 SO Sequence 429 AA:

Query Match 10.0%; Score 651.5; DB 21; Length 429;
 Best Local Similarity 36.9%; Pred. No. 1.2e-23;
 Matches 165; Conservative 48; Mismatches 169; Indels 65; Gaps 15;

QY 813 GLAAG-----TAYTVVAANDAGNTSAPST---PYTANT-----TSPS-P-849
 DB 13 gllaalgavlaafvssvsaapagadtlcepfgtllgryvvnrtwsgtapgcv 72
 QY 850 SPTPTGTTTDCPTGPNONGV---TSVGDDEXRYOTNEMNSAOOCLTINTATGAWTS 905
 DB 73 laldgfrtqdgdsaprtgarkyspvtngchylqnmwsgtstapcvta-ldtg-frvt-130
 QY 906 TANFSGGTGAPATYPSIYKGMNCTTKNQMPIOISGASVSWSTTQVSSGAYDV 965
 DB 131 qdgdsaprtgarkyspvtngchylcnc-spgtdlpyrlctvsaspslsygfvdgavyna 189
 QY 966 AYDIWTNSTPTTGGOPNTEIMTLNRSKGVOPFGSQTATGTVACHTWNVMWOGQOTSMK 1025
 DB 190 sydlwldprlartldg-vnqetelmwlnrvyrlqprvglt-asvgrtlevwsgngsgnd 247
 QY 1026 IISYVLTPTATSIINLDAIFADAARGLNSTDYLDVEAGFEIMOGGOGGSNSFVS 1085
 DB 248 vstsv-apalsgwsfdvmdfratvarjlaendwyltsvqagfepwqngaglaavsfss 306
 QY 1086 SVTSGTSSPTSPSPPTPTSPSPPTSPSPSSSGVACRATVYVNSDMGSGF 1145
 DB 307 tveltgpggt-----dpbgdpgps-----acavsygltv-wqdgf 340
 QY 1146 TAYTVVNTNGSRATNGVYAMSEFGNQTYNTWNTALITOSGASVTATNLSTYNNVIOPGOS 1205
 DB 341 ladvltvntngtldpvdwqlaflpsqgtiltnawnaaltpssgvtatgshnarlapgs 400
 QY 1206 TTEGFNGSYSGTNAAPT-----LSCT 1226
 DB 401 lsfqfgylygafaeptcflngtact 427

RESULT 9
 AAY67497
 ID AAY67497 standard; Protein; 386 AA.
 XX AAY67497;
 XX
 DT 19-MAY-2000 (first entry)
 XX
 DE Protein sequence of Cella and cellulase 11A68 fusion.
 XX Cellulase; Actinomyces; detergent; feed additive; textile treatment;
 KW

KW pulp; paper; cellulase 11A68.
 XX Streptomyces lividans.
 OS Synthetic.
 XX

EH Key Location/Qualifiers
 FT Peptide 1..46
 FT Protein /note="Cela signal sequence"
 FT /note="cellulase 11A68 mature peptide"

WO200009707-A1.

24-FEB-2000.

28-MAY-1999; 99WO-US11971.

24-JUN-1998; 98US-0104308.

18-NOV-1998; 98WO-US24649.

28-MAY-1999; 99US-0321981.

(GENV) GENENCOR INT INC.

Jones BE, Van Der Kleijf WAH, Van Solingen P, Weyler W;

WPI; 2000-224344/19.

N-PSDB; AA57031.

A novel Actinomyces cellulase and related DNA, useful for detergent compositions, treating cellulose and paper or pulp -
 Example 6; Fig 15; 72pp; English.

The invention provides a cellulase from Actinomyces. The cellulase can be used in a detergent composition, as an additive for animal feed and for the treatment of textiles or pulp and paper. The DNA encoding the cellulase can be used to identify homologous cellulases and for recombinant production of cellulases. The present sequence represents the protein sequence of a cela signal sequence and cellulase 11A68 GI promoter, cela signal sequence, cellulase 11A68 and GI terminator

Sequence 386 AA:

Query Match 9.5%; Score 622.5; DB 21; Length 386;
 Best Local Similarity 38.6%; Pred. No. 2.7e-22;
 Matches 152; Conservative 51; Mismatches 132; Indels 59; Gaps 18;

QY 854 TGTVTIDCT--PQP-----GVTSGDDEXRYOTNEMNSAOOCLTINTATGAW 902
 DB 29 tgytslvgltalalappagangqldrygttllg-dryvqnmrtvgtaatgc--lnvtngnft 85
 QY 903 TVSTANFSGGTGAPATYPSIYKGMNCTTKNVMPIOISGASVSWSTTQVSSGA 962
 DB 86 eltqdgsvpvtngarkyspvydgchngncaprtl-lpmrlsisaapsavsyfrylcnv 144
 QY 963 YDVAAYDIWTNSTPTTGGOPNTEIMTLNRSKGVOPFGSQTATGTVACHTWNVMWOGQOT 1022
 DB 145 ynaaydwlldprlmg-varlelmwlnrvpvpqprgspvglt-ahvggrfswewtqsgn 202
 QY 1023 SMKIIISYVLTPTATSIINLDAIFADAARGLNSTD-YLLDVEAGFEIMOGGOGGSN 1081
 DB 203 sndvlsf-lapsalswsvfvdg-fvdgavshjlatpbdwyltslqngfepwqngtqlavn 260
 QY 1082 SFSVSVTS-GTSSPTSPSPPTPTSPSPPTSPSPSSSGVACRATVYVNSD 1140
 DB 261 sfssavnaaggngstpg-----tpa-----acqvsyecht- 290
 QY 1141 WSGGFTATVTVNTNGSRATNGVYAMSEFGNQTYNTWNTALITOSGASVTATNLSTY-NV 1199
 DB 291 wpgsfvtdtlntngtldpvdgweldflpagnhtvtaamallspaagavlatarstgngrl 350

OY 1200 IQPGSTFGNGSYSGTN-AAPT-----LSCF 1226
 Db 351 aangtgcgfgtgcstgtnapaggringtsct 384

RESULT 10
 AAY06367
 ID AAY06367 standard; Protein: 371 AA.

AC AAY06367;

06-SEP-1999 (first entry)

Streptomyces sp. EgIII-like cellulase.

Cellulase; endoglucanase; EgIII; textile; feed additive; baking;
 food processing; grain wet milling; pulp; paper.

Streptomyces sp.

MO9931255-A2.

24-JUN-1999.

14-DEC-1998; 98WO-US26552.

16-DEC-1997; 97US-0991720.

(GENEV) GENENCOR INT INC.

Bower BS, Fowler T, Phillips JT;

WPI; 1999-395187/33.

EgIII like cellulase

Example; Fig 6; 47pp; English.

The present polypeptide represents a full-length sequence of a novel EgIII-like cellulase of Streptomyces sp. 11A8. It was deduced from a gene sequence isolated from genomic DNA using PCR primers (see AAY59180-91) based on conserved motifs (see AAY06325-29) of Trichoderma reesei EgIII cellulase and related enzymes. PCR has been used to identify novel EgIII-like enzymes, including the present protein, from bacterial and fungal sources (see AAY06331-70). The sequence shows homology to T. reesei EgIII (see AAY06330). Also provided by the invention are vectors, host cells and methods for the recombinant production of such enzymes, which can be used in the treatment of cellulose-containing textiles, as feed additives, in the treatment of wood pulp, in the reduction of biomass to glucose, in the stone washing of indigo dyed denim, or as laundry detergent components (all claimed).

Sequence 371 AA;

Query Match 9.5%; Score 620; DB 20; Length 371;
 Best Local Similarity 38.7%; Pred. No. 3.4e-22;
 Matches 150; Conservative 50; Mismatches 134; Indels 54; Gaps 17;

OY 855 GTTVDCTPG-PNON-----GVTSGVQGEYRYQTENMNSSAOQCLTINTAGAWTVSTAN 908
 Db 20 galltaapagangqldrygtlttq-dryvvqmrwgtstqtc--lntvngfcltqad 76

OY 909 FSGGTGAPATYPSITKCGHNGNCTKNVGNPDIQISGSAVTSMTQVSSGAYDAVD 968
 Db 77 gsvptngapkeypsvydgcchgnacprtt-lpmrtsslgaspsavsytytngyynaayd 135

OY 969 IWTNSTPTTQDPNGTEITMILNSRGVQPGSGTATGVTYAGHTWNVQGOQTSWKIIS 1028
 Db 136 lwdlptptng-vnrteimifnrvygpjgspygt-ahvgzsewewtsgnsndvys 193

OY 1029 YVLPGATISINLDKAFADAAARGLNTSD-YLDDVEAFETIWMGGOGIGLSNFSVS 1087

Db 194 f-lapsaaiswefdvkd-fvdqavghjcltpwyltsiqagtepwegtlavnsfssav 251
 OY 1088 TS-GTSSPTSPSPSPPTPSPSPSPSPSPSPSSSSGACRATYVNVSDMGSGFT 1146
 Db 252 nagsgnggtlp-----tpa-----acqvstltc-wpggft 281

OY 1147 ATVTYNTNGSRATNGWYVAMSFGNQTYNTYNTNTALTOGASVYATNTN.SYN-NV1QPGOS 1205
 Db 282 vdlitntcgtstpvdgweldfltpaghtlytsawnalispasgavtarsgnsgrlaangft 341

OY 1206 TTFGFGNGSYSGTN-AAPT-----LSCF 1226
 Db 342 gsfgtglsstgtnapaggringtsct 369

RESULT 11
 AAB14880
 ID AAB14880 standard; Protein: 371 AA.

AC AAB14880;

21-NOV-2000 (first entry)

Streptomyces lividans CelB EgIII-like cellulase.

Streptomyces lividans; Trichoderma reesei; endoglucanase III; EgIII;
 cellulase; mutant; enzyme stability; textile treatment;
 wood pulp treatment; feed additive; detergent.

Streptomyces lividans.

WO200037614-A2.

29-JUN-2000.

12-NOV-1999; 99WO-US26704.

18-DEC-1998; 98US-0216295.

(GENEV) GENENCOR INT INC.

Mitchinson C, Wendt DJ;

WPI; 2000-482483/42.

Novel endoglucanase III or endoglucanase III-like cellulase useful for treating textiles and wood pulp comprises a substitution or deletion at specified positions in the wild form of endoglucanase III -
 Example 1; Fig 3; 52pp; English.

The present sequence is a cellulase related to endoglucanase III (EgIII) from Trichoderma reesei. EgIII-like genes were isolated from genomic DNA libraries constructed from various microorganisms by PCR. The isolated genes showed significant homology to EgIII from T. reesei. Certain substitution and deletion mutations have been incorporated into EgIII and EgIII-like cellulases to produce variant enzymes with improved stability, e.g. increased resistance to temperature stress. The mutants may be used in textile and wood pulp treatment, as a feed additive, and for reducing biomass to glucose. They are also useful for stonewashing or indigo dyed denim and as an agent in laundry and dish detergents.

Sequence 371 AA;

Query Match 9.5%; Score 620; DB 21; Length 371;
 Best Local Similarity 38.7%; Pred. No. 3.4e-22;
 Matches 150; Conservative 50; Mismatches 134; Indels 54; Gaps 17;

OY 855 GTTVDCTPG-PNON-----GVTSGVQGEYRYQTENMNSSAOQCLTINTAGAWTVSTAN 908
 Db 20 galltaapagangqldrygtlttq-dryvvqmrwgtstqtc--lntvngfcltqad 76

QY 909 FSGGTGAPATPYISYKGCWMCKTNNKGMPIQISOIGSAVTSWSTTQVSSGAYDVAYD 968
 Db 77 gsvptngapksypsvydgchyncaprtt-lpmrissigsapsvsyrytgnvyanaad 135
 QY 969 IWTNSTPTTTGQPNGTETIMLWNSRGVOPFGSQTATGTVYAGHTMVMVNGOQTSKIIIS 1028
 Db 136 Iwldprrtng-vnrteimfwfnrvpvgqjgspvgt-ahvgrswewwvsgnsndvis 193
 QY 1029 YVLTGATSTISNLDKAIIFADAARGLANTSD-YLIDVEAGFEIMOGGGLGSNFSVSV 1087
 Db 194 f-lapsaiswsfdvkd-fvdqavshglapdwylltsigqgfepwgggtclavnsfsasv 251
 QY 1088 TS-GTSSPTPSPSPPTPPSPPTPPSPSPPTSSPSSSGVACRAITYVNSDMWGSQFT 1146
 Db 252 naqgnggtpg-----tpa-----acqvsyatcht-wpggft 281
 QY 1147 ATVTYNTGSRATNGWTVAMSFEGNQTVTNYMTALTGSGASTATNLSTN-NVIOPGQS 1205
 Db 282 vdtltntgstpvdgweldltlpaghtvsaamallspasgavtarstgngriaangt 341
 QY 1206 TTEFGNYSYGTN-AAPT-----LSCT 1226
 Db 342 qsfqfgtsasgtgfnapaggrlntgact 369

RESULT 12

AA84345
 ID AAY84345 standard; Protein; 371 AA.

AC AAY84345;

DT 12-JUL-2000 (first entry)

DE Amino acid sequence of an endoglucanase III (EGIII)-like cellulase.

KW Endoglucanase III; EGIII; EGIII-like cellulase; surfactant stability;

KW cellulase; textile processing; textile cleaning; stonewashing;

KW indigo dyed denim; cellulose containing fabric; fabric smoothness;

KW pill removal; fibril removal; cotton; cellulosic fibre; dyeing; detergent;

KW animal feed; wood pulp; paper; grain; biomass reduction; glucose.

XX Actinomyces sp.

OS WO200014208-A1.

PN 16-MAR-2000.

PD 24-AUG-1999; 99WO-US19154.

PE 03-SEP-1998; 98US-0146729.

PR (GENEV) GENENCOR INT INC.

XX Fowler T;

PI MPI; 2000-271052/23.

DR Novel variant endoglucanase III-like cellulases with improved

XX surfactant stability and resistance to temperature stress, useful for

PT textile processing or cleaning, treating wood pulp, food and grain, and

XX reducing biomass to glucose

PS Disclosure: Page 64-65; 73pp; English.

XX The present sequence represents an endoglucanase III (EGIII)-like

CC cellulase. The cellulase has homology to the Trichoderma reesei EGIII

CC protein. The variant cellulases have improved temperature stability,

CC and improved surfactant stability. The variant cellulases and

CC compositions containing them are used in textile processing or cleaning,

CC e.g. stonewashing of indigo dyed denim, and modifying the texture, feel

CC or appearance of cellulose containing fabrics (e.g. improving fabric

CC smoothness or removing pills and fibrils). The compositions may also be

CC used for the removal of immature or dead cotton from cellulosic fibres

CC or fabric, which can cause uneven dying. The cellulase may also be used

CC in a detergent composition for washing laundry and dishes and in the

CC treatment of animal feed, wood pulp, paper, non-animal foods and grains.

CC The enzymes may also be used in the reduction of biomass to glucose.

XX Sequence 371 AA;

QY Query Match 9.5%; Score 620; DB 21; Length 371;

Db Best Local Similarity 38.7%; Pred. No. 3.4e-22;

Matches 150; Conservative 50; Mismatches 134; Indels 54; Gaps 17;

QY 855 GTTWTDCPTG-PKON-----GVTSVQDEXRVQETNENSSAOCLTINTAGMTVSTAN 908

Db 20 galltaapaganqldrdygtltlg-dryvqgnrvgtsatgc--lnvqngfelqad 76

QY 909 FSGGTGAPATPYISYKGCWMCKTNNKGMPIQISOIGSAVTSWSTTQVSSGAYDVAYD 968

Db 77 gsvptngapksypsvydgchyncaprtt-lpmrissigsapsvsyrytgnvyanaad 135

QY 969 IWTNSTPTTTGQPNGTETIMLWNSRGVOPFGSQTATGTVYAGHTMVMVNGOQTSKIIIS 1028

Db 136 Iwldprrtng-vnrteimfwfnrvpvgqjgspvgt-ahvgrswewwvsgnsndvis 193

QY 1029 YVLTGATSTISNLDKAIIFADAARGLANTSD-YLIDVEAGFEIMOGGGLGSNFSVSV 1087

Db 194 f-lapsaiswsfdvkd-fvdqavshglapdwylltsigqgfepwgggtclavnsfsasv 251

QY 1088 TS-GTSSPTPSPSPPTPPSPPTPPSPSPPTSSPSSSGVACRAITYVNSDMWGSQFT 1146

Db 252 naqgnggtpg-----tpa-----acqvsyatcht-wpggft 281

QY 1147 ATVTYNTGSRATNGWTVAMSFEGNQTVTNYMTALTGSGASTATNLSTN-NVIOPGQS 1205

Db 282 vdtltntgstpvdgweldltlpaghtvsaamallspasgavtarstgngriaangt 341

QY 1206 TTEFGNYSYGTN-AAPT-----LSCT 1226

Db 342 qsfqfgtsasgtgfnapaggrlntgact 369

XX Actinomyces sp.

OS WO9925846-A2.

PN 27-MAY-1999.

PD 18-NOV-1998; 98WO-US24649.

PE 24-JUN-1998; 98US-0104308.

PR 19-NOV-1997; 97US-0974041.

XX 19-NOV-1997; 97US-0974042.

XX (GENEV) GENENCOR INT INC.

XX Jones BE, Van Der Kleij WAH, Van Solingen P, Weyler W;

XX

PI

...

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:18:25 ; Search time 72.78 Seconds

(Without alignments)
1621.293 Million cell updates/sec

Title: US-09-917-384-1

Perfect score: 6525

Sequence: 1 MERTQOSGRNCRYOGRTRM.....GFNGSYSGTNAAPTLCSTAS 1228

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 100%

Listing first 45 summaries

Database : PIR71:*

1: pirl:*

2: pirl:*

3: pirl:*

4: pirl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1655.5	25.4	872	2	cellulase - Cellul
2	1358	20.8	549541	2	1,4-beta-cellulobios
3	1051	16.1	579	2	probable secreted
4	1000	15.3	596	2	cellulase 1,4-beta
5	744.5	11.4	1742	2	cellulase (EC 3.2.
6	743.5	11.4	1711	2	1,4-beta-glucanase
7	675	10.3	1779	2	xylanase - Caldice
8	671.5	10.3	1331	2	mannan endo-1,4-be
9	649.5	10.0	382	2	cellulase (EC 3.2.
10	641	9.8	438	1	cellulose 1,4-beta
11	581.5	8.9	1090	2	cellulase 1,4-beta
12	580	8.9	1045	2	endoglucanase B (E
13	550	8.4	2232	2	hypothetical prote
14	548	8.4	471	1	cellulose 1,4-beta
15	548	8.4	471	1	cellulose 1,4-beta
16	519.5	8.0	1039	2	cellulase (EC 3.2.
17	505	7.7	747	2	cellulase (EC 3.2.
18	488	7.5	1854	2	cellulase (EC 3.2.
19	473.5	7.3	473	2	clpA protein - Clo
20	471	7.2	3020	2	hypothetical prote
21	443	6.8	1367	1	mucin 2 precursor,
22	416.5	6.4	474	2	glucan 1,4-alpha-g
23	411	6.3	279	2	protein TPX-VT3 -
24	405.5	6.2	915	2	hypothetical prote
25	404.5	6.2	1275	2	cellulase (EC 3.2.
26	404	6.2	441	2	hypothetical prote
27	398.5	6.1	1032	2	cellulase (EC 3.2.
28	393.5	6.0	4776	2	hypothetical prote
29	387.5	5.9	1365	2	cell wall surface
					hypothetical prote

30	386	5.9	145	2	cellulase homolog
31	385.5	5.9	1777	2	hypothetical prote
32	382.5	5.9	3570	2	mucin MUC5B, trach
33	379	5.8	1459	2	hypothetical prote
34	378	5.8	1459	2	hypothetical prote
35	377.5	5.8	700	2	cellulase (EC 3.2.
36	375.5	5.8	499	2	cellulase (EC 3.2.
37	375	5.7	1483	2	cellulase (EC 3.2.
38	374.5	5.7	351	2	probably celluloso
39	370.5	5.7	508	2	hypothetical prote
40	369.5	5.7	1664	2	cellulase (EC 3.2.
41	369	5.7	890	2	S-layer protein
42	366	5.6	499	2	probable secreted
43	363.5	5.6	2271	2	cellulase (EC 3.2.
44	359	5.5	1104	2	hypothetical prote
45	355.5	5.4	380	2	probable membrane
					probable cellulase

ALIGNMENTS

RESULT 1

S49541

cellulase - Cellulomonas flm1

C:Species: Cellulomonas flm1

C:Date: 29-Nov-1995 #sequence-revision 01-Mar-1996 #text_change 22-Oct-1999

C:Accession: S49541; A47093

R:Meinke, A.; Gilkes, N.R.; Kwon, E.; Kilburn, D.G.; Warren, R.A.J.; Miller Jr., R.C.

Mol. Microbiol. 12, 413-422, 1994

A:Title: Cellulohydrolase A (Cbha) from the cellulolytic bacterium Cellulomonas flm1

A:Reference number: S49541; MUID:94344030

A:Accession: S49541

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-872 <ME1>

A:Cross-references: EMBL:L25809; NID:9456028; PIDN:MAC36898.1; PID:9456029

R:Meinke, A.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.

J. Bacteriol. 175, 1910-1918, 1993

A:Title: Cellulose-binding polypeptides from Cellulomonas flm1: endoglucanase D (Cend

A:Reference number: A47093; MUID:93209933

A:Accession: A47093

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 812-872 <ME2>

A:Experimental source: ATCC 484

A:Note: sequence extracted from NCBI backbone (NCBI:128120, NCBI:128121)

C:Superfamily: bacterial cellulose-binding domain homology

F:769-870/Domain: bacterial cellulose-binding domain homology <BCB>

F:770-869/Disulfide bonds: #status predicted

Query Match

Best Local Similarity 33.18; Pred. No. 2,4e-62; Indels 399; Gaps 28;

Matches 408; Conservative 125; Mismatches 300;

Query 23 ISKRLRAGVLAGAVASIAIVPLA-MQHPAIAITHVNDPAGATFFVNPMAQEVQSEA 80

Db 10 VRRVRA-VATAATATLVANPLTTATSAVAHVNDPAGAVQVYVNPMAASVNA 68

Query 81 ANQT-AMTAAKRVSTYSTAVMDRIAINGVNGGGLTTYDAALISQOQGT-TPEVI 138

Db 69 GROSADPALAKMRTVAGOPAVMMDRISATIGNADNGSLFHDNVAQOKAAGVPLVF 128

Query 139 EIVYDIPGRCAALASNGELPATRAGQETEQYIDPIASITLNPYSSRIYTIIEPD 198

Db 129 NLVYDIPGRCAALASNGELPATRAGQETEQYIDPIASITLNPYSSRIYTIIEPD 188

Query 199 SLPAVATNMSIOACATAVPEYEGIEYALTKLHAIIPVYIYMDAHSGLMPNNA--S 255

Db 189 SLPLUTINISBPACQAPARYRGQVKYALDKLHAIIPVYIYMDAHSGLMPNNA--S 248

Query 256 GYQEVQKVLNASTGVNGIDGFTVNTAYTPPLAKPEPT-ATQVGGQPVESANFYQWNP 314

Db 256 GYQEVQKVLNASTGVNGIDGFTVNTAYTPPLAKPEPT-ATQVGGQPVESANFYQWNP 314

Db 249 TLFAYAK--STTAGFASIDGFVSDVANTTPLEEPLSDSSSLTINNTPIRSSKFEYMNFD 306
QY 315 IDEAYAVDLRYLVAAGPSSSGMIDLRKMGKGNPTGPSTTDVNTVFNOSKIDL 374
Db 307 FEIDITAHMRLVAAGPSSSGMIDLRKMGKGNPTGPSTTDVNTVFNOSKIDL 366
QY 375 ROHRLMGNQAGLQGPQASPTDPPNNAHLDAYVYIKPGSGSDGSAASDPTTGKSP 434
Db 367 RYHRGAKMCPDLAGICORFEATPSSGAASHLDAFWIKPGSGSDGSDTIPDQGRFR 426
QY 433 MCDPITYTS--YGVLTNALPNSPIAGQWPAQFDOLVANARAPVPTSTSSPPPPSPS 492
Db 427 MCDPFEVSKLNNQLTGAPPNAPLAGQWFEQFVTLVKNAPVIGGT-----FVEDL 479
QY 493 ASPSPSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSS 552
Db 480 VAPT----- 483
QY 553 PSPSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSS 612
Db 484 -----VPTGLTAGT 492
QY 613 TSSSSVDLSTVVRKWFTRDGSSTLVYNCNMAAGCGRIRASFGSVNPAFTADTYQL 672
Db 493 TTAATSVPLS-----WTA-----STDNAAVGYDYVRCGT 520
QY 673 SFTGGTAGSGTGEIONVKNKSDMSNFTETNDYSGTNTFQDMKVVYVNGVYMGCT 732
Db 521 TLVGTTAA-----TSYTVTGLTPATVAY--SFTVRK 549
QY 733 EPSGSP 789
Db 550 DAAGNVAASAAATTTGGGTVDTPAPSVPAGLTAGTTTTTTPSLWASIDNMGSGV 609
QY 790 AATYTRKRVLVGPVTSFTDTGLAGTAITYTAAVDAAGTSAFSPVAT----- 843
Db 610 AGYEVLKRGTTVGTATNISTYVGLTAGTTSFSAKADVAANTSAAVAASATQGTGV 669
QY 844 --TSPS--PSPPTGTTVDCTPGPNQNGVTSVQDEYVQINEMMSAQQCLITATAG 900
Db 670 VTTAPSVPTGLTAGTTTTSSVP-----LT----- 694
QY 901 AATVSTANFSGGTGAPATYPSIYKCHWNGCTKNVGMPIQISQISAVTSMTQVSS 960
Db 695 --WTASIDN--AGSGVA----- 708
QY 961 GAYDAVYDWTSTPTTGTQPNGTETMLWLSRGGVQPFQSGOATGVVAGHTWYMGQ 1020
Db 709 -----GYEYFNCTTTRVAT-----VTSSTYTVTG----- 731
QY 1021 OTSMKISIVLTGATISINLDKAIFADAAARGLSNTSDYLLDVAGFEINOGGGLGS 1080
Db 732 -----LAADTAISFTYKADVANVASA----- 754
QY 1081 NSFVSATSGTSPS 1140
Db 755 ---SAAVASRGTQAT-----SG--GCYIKYKASS- 778
QY 1141 WSGGTAATVYVNTGSRATNGTVAASFGCNOTVTYNTALTATOGASVYATINLSYNNV 1200
Db 779 NMTGFGTGVVKNNGTAAALNGWTLGFSFADGOKVSGQMSAEMWSQSTAVTAKNAPANGTL 838
QY 1201 OPGOSTTGFNGSYSGTNAAPT-----LSC 1226
Db 839 AAGSSYSIGFNHGTNTAPTAFTLNGVACT 870

RESULT 2
AB2704
1.4-beta-cellobiosidase XF1267 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: AB2704

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: AB2515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: AB2704
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-683 <SIM>
A:Cross-References: GB:AE003960; GB:AE003849; NID:g9106242; PIDN:AAF84076.1; GSPDB:GN
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Brienes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrero, D.M.; Carre
as Neto, E.; Docena, C.; El-Dorri, H.; Facincanli, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Klajima, J.P.; Krieger, J.E.; Kurama, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, R.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, C.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovsky-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1267

Query Match 20.8%; Score 1358; DB 2; Length 683;
Best Local Similarity 43.9%; Pred. No. 5.2e-50;
Matches 319; Conservative 81; Mismatches 218; Indels 108; Gaps 19;

QY 53 AATHDNPYAGATFEVNPYWAQEVQSEANQNTATLAARKRVYSTVAMDRRAAING 112
Db 18 AEAHNDPEVDATSYLNPDIYSKIDISIAKVDVSLKEKMQTKIKNPYGVMTDITDAVG 77
QY 113 VNGGPG---LFTYTLDAALSQOQGTPEVEIYIYDLPGDCALASNGELPATAGLQTY 169
Db 78 GPRNPERLNLGTGLDAALQKKGMPITATFYIYDIPGDCALASNGELPATAGLQTY 137
QY 170 ETQYIDPIASILNPKYSSRLIYTIIEPDLPAVYNNMSIQACATA--VPYEOGLEYVL 227
Db 138 KREYIDTASISNPKYKRIYNTIIEPDLNITNMTSPSCAQAKNGIYEEGKIKAL 197
QY 228 TKLHAIPNYIYMDAASGWLGMPPNNAAGYVOEVQVL--NASIGVNGIDGYTNTANTYP 286
Db 198 NKLSEIPNYNYNMDDIGHSGWLGMIDTNRIPAVSLYTKVIGSTAGFASVNGPATNTANTYP 257
QY 287 LKEPFM--TATQOVGCGPVASANTYQWNPIDEDVAVDIXSRLYAAGPSSIGMLIDTLR 345
Db 258 LIEPNLPNDLNLGGQPIRSSKFEYENRYFDEMDYSELYNDFVAAAGWPSISIGFIIDGR 317
QY 346 NGMGEPNEPTGPSTATDVNTFVNOBKIDLRHRLMGNQAGLQGPQASPTDPPNNAH 405
Db 318 NGMGEPNPTS--AFGNDVANSYNSGRIDRRNRHGMKCNKQKAGIGLPIAT-----PGCHV 372
QY 406 DAYVWIKPGESDGTSAASDPTTGKKSDDPCTTYTTSYGVLTNALPNSPIAGQWPAQF 465
Db 373 DAFOWKIPPGYSGSSSLIPDQKGFDRYCDPTFTPGVLTGALPDLPSGDWHAQF 432
QY 466 DQLVANARAPVPTSTSSPPPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 525
Db 433 VKLINNAYPD-----SKTPPLGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 486
QY 526 SPSSSP 585
Db 487 SGSS 546
QY 586 GGLKLYQKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVYRYWFTTRDGSSTLVYNCDMA 645
Db 547 SG-----SSSSSSSSSS-----SSGSSS-----GSSSGSGFN----- 572

Query Match	Similarity	16.1%:	Score: 1051:	DB 2:	Length 579:
Best Local Similarity	49.2%:	Pred. No. 3.1e-37:			
Matches 215:	Conservative	61:	Mismatches 131:	Indels 30:	Gaps 8:
Qy	50	PAIAATHVNDNPYAGATFFVNPYMAQEOVSANQTNATLAAKMRVSYSTSTAVWMDRIA	109		
Db	158	PTGPGDVRDNPFDYGAQVYVNPMSANNAAPGGR-----IADEPGVWLDRIA	207		
Qy	110	INGVGGGGLTTLDAALSSQOQGTPEYIEYITDLPGRCCAAALNSGELPTAAGLOTY	169		
Db	208	IEGANGGGLRDLDAAL-EDQSGSEMYVQVLYLPRDSCSALASNGELGPT--IDRY	264		
Qy	170	ETQYIDPLASILSNPKYSRLVITIEPDSLPTNATVNSIQACAT-----AVPYEKG	222		
Db	265	KTEYIDPLAEILSDSKYADLRITVETLDSLPNLVTNYSGRPTAENDYKAKANNYOKG	324		
Qy	223	IEYALTKLHAIPNVYIYMDAAHSGWLGMPPNNAAGVYQVOKVLNASIG--VNGIDGFEYTN	280		
Db	325	VGYALNRGAAGVNVYNYVDAGHHGWLGMDSN-FGSPAEIFKTAATTEGATLDDVYGFIVN	383		
Qy	281	TANYPLKEPFMTATQOVGGQRPVESANTYQWNPDIDEDYANDLXSLVLAAGFSPSITML	340		
Db	384	TANYALKEENFKITDSVNGTSVROSDVWDMQYIDELSTYQAMADKVLVSIGFDNLML	443		
Qy	341	IDTLRNGGGRPEPTSTATDNTFVNSQKIDLRQHRGLMCNONGAGIGLOPQASPTDF	400		
Db	444	IDTSNCGMGADRPFGPGATITDVTNYYVNGGRDRIRHLGMCNONGAGIGERQASPA--	501		
Qy	401	PNHALDAYVWIKRPESDQTSAAASDPPTTGKSSDPCDEPYT--TSYGLVINALPNSPIA	457		
Db	502	--AGIDAYWMPKRPESDQSSKTLIDNPDGKGFDRMCDPLYTGNERNGNMSGALPDAPIS	559		
Qy	458	GOMFPAQDFQVLANARP	474		
Db	560	GAMFSAQGFQELKKNAP	576		
RESULT	4				
A55976					
cellulose 1,4-beta-cellulohydrolase (EC 3.2.1.91) - Thermomonospora fusca					
C:Species: Thermomonospora fusca					

[illegible]

RESULT 5
T17120
cellulase (EC 3.2.1.-) precursor, thermoactive - *Caldocellum saccharolyticum*
C.Species: *Caldocellum saccharolyticum*
C.Date: 15-Oct-1999 #sequence=revision 15-Oct-1999 #text-change 17-Nov-2000
C.Accession: T17120; A43745
R.Rev: V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A.Title: CelA, another gene coding for a multidomain cellulase from the extreme thermophile
A.Reference number: Z18658; MUID:95336703
A.Accession: T17120
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-1742 <REO>
A.Cross-references: EMBL:L32742; NID:G537499; PID:5537500; PIDN:AA91086.1
R.Uethli, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A.Title: Cloning, sequence analysis, and expression in *Escherichia coli* of a gene cod


```

OY 930 -----GNCITKNYGMFIQ-----ISOISAVT-----SMSTVOSSGAYDAVDIMTN 972
Db 1481 FGGVPGQDAMPNTARVLPQDDMISVYGSATIPADYMEGDSGINKMTYYSYCTN 1540
OY 973 STPTTGGPNGTEIMWILNSRGVOPFGSOTATGVT-----VAGHT----- 1013
Db 1541 FAQRGSGFPAGALIVMT-----GKSPVPMWEKRGVILNPNGFVGVGNHHLQLEFNGK 1596
OY 1014 WNWOGQOTSMKIIISVLTPEGATS-----ISNLDKAIKPADAAKSLNTSDYLLDVE 1066
Db 1597 WYIYVHAQTKKLDLG--VAKGRSPHINQVOIENGVIKEVTADYKGAQVKKFDPRMVE 1654
OY 1067 AGEIWMOGS 1075
Db 1655 AETPAMCAG 1663

RESULT 8
A48954
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum
N:Alternate names: beta-mannanase
C:Species: Caldocellum saccharolyticum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A48954; B43745
R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain
A:Reference number: A48954; MUID:93119139
A:Accession: A48954
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1331 <GB>
A:Cross-references: GB:L01257; NID:g144290; PIDN:AA71887.1; PID:g144291
A>Note: sequence extracted from NCBI backbone (NCBIN:121576, NCBI:P:121577)
R:Luehth, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding
A:Reference number: A43745; MUID:91247819
A:Accession: B43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337, 'PROHQRHQR' <LUE>
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AA72861.1; PID:g144294
A>Note: this sequence has been revised in reference A48954
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 10.3%; Score 671.5; DB 2: Length 1331;
Best Local Similarity 23.5%; Pred. No. 6.3e-21;
Matches 243; Conservative 145; Mismatches 305; Indels 341; Gaps 43;

OY 240 MDAHSGMLGWPNN-----ASGY-----VOEVQKV--LNASIGVNGI----- 274
Db 67 LDTALRGIRSMQMSNVRVLVLSNGYRWTKIPASEVANITLSLSLGRKATILEVHDPTGVC 126
OY 275 -DGFPTNAN-----YTPLK-----EPFMTAQOYGGQVSEANATQNPDIIDEADYAVD 323
Db 127 EGGACSLAQAVEYWEKRSVLDGNEDEYIT--NNGNEFYGNNNYONWVNDTKNAIKA-- 182
OY 324 IYSRLVAAGFPSSIGMLIDITLRNG--WGGPNEPTGPS-----TA 360
Db 183 -----LRDAGFKFTI--AMDAPWQGDWSTMTMDNQSITHEADPLRLNLYFSIMTYGYNTA 236
OY 361 TDVNTFVNQSKIDLROHRLMCNONGAGIGQPPQASPT--DFPNAHLDAYVWIKPPGSS 417
Db 237 SKVEYFI-KSFYD-----KGL-----PLVIGFEGHNDG-----DPDE 270
OY 418 DGTSAASDPTTKKSDPMC-DFTYTTSGVLTNALPNSPIA-GQWPPAQFDOLVANARPA 475
Db 271 AIVRYAKQYKILIFSWMSMGNSGYGLDMVANNMDPNNTPPMGQWYKTN-----A 320
OY 476 VPTSTSSPPPPPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 535

```

```

Db 321 IGT----- 331
OY 536 SSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 594
Db 332 TYPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 373
OY 595 NDSAPGNOIKFGLQLVNTGSSVDLSYTVRYWFTTRDGSSTLVYNCDMAAGCGNIRA 654
Db 374 KETNSTTWTIRFWLKVYVNGSSSIDLSRVTIRYWTVVGERSAAS--DMAQIGASNTVF 432
OY 655 SFGSVNPAPTPTADTYLOLSTFGCT--LAAGSTGSIORVYKNSDMNSNTEFTDYSY-GTN 711
Db 433 KTVKLSSVSGADYTLLEGFKSGAGOLQPKDTGELQMFNFNDMSNTNQGDMWSIQSM 492
OY 712 TTPQDMTKVTVVNGVLVWGTGTSPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 771
Db 493 TSYGENEKYTAIDGVLVWQGPSCGATAPATATPTPTPTPTPTPTPTPTPTPTPTPTPT 536
OY 772 SGSSVSLMNASTDVNGVAHVYRNGVLVQGPYTSFTDTGLAAGIATYTVAAVDAAG 831
Db 537 -----TPPTV----- 541
OY 832 NTSAPSTPTATTTSPSPSPPTGTVTYDCTGPNQNGVTSVQGDERY-----QTNEMN 886
Db 542 -----PTPTVTAT--PTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 583
OY 887 SSAQOL-TINTATGAMVYSTANFSGGTGCGAPATYPSIKGCHMGNCCTTKVGMPIQISQ 945
Db 584 NTRPMLKLVNNGSSSIDLSRVTIR-----YV-----YVDDGR 617
OY 946 TCSAVTSMSTVOSSGADVAVDITNSTPTTGGPNGTEIMWILNSRGVOPFGSOTAT 1005
Db 618 AQSALSDMA---QIGASVWTFK-FVKLSSSVSGADYVLEIG-FKSGAGOLQPKDGTGRI 671
OY 1006 GTVAVGHTM-VNWOGQOTSMKIIISVLTPEGATSINLDKAIKPADAAKSLNTSDYLLD 1064
Db 672 QIRFKSDMSNINQGNDSW-----IQSMTSYGENEKYTVATID----- 709
OY 1065 VEAGEIWMOGGGLGSLNSPFSVTSQTSRSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1124
Db 710 ---GVLVW-GQPSGCTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 765
OY 1125 SSSGVACRATTVYVNSDMSGFTATYTV-TNTGSRATNGTAVMSGNGQCTVNTNTALT 1183
Db 766 TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 821
OY 1184 QSGASVTATNLISYN 1197
Db 822 LQGNRLITGYNMENN 835

RESULT 9
JC2571
cellulase (EC 3.2.1.4) precursor - Streptomyces rochei (strain A2)
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase
C:Species: Streptomyces rochei
C:Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 22-Oct-1999
C:Accession: JC2571; S34392
R:Perillo, B.; Hanhart, E.; Irdani, T.; Iqbal, M.; McCarthy, A.J.; Mastromel, G.
A:Title: Characterization and sequence analysis of a Streptomyces rochei A2 endoglucanase
A:Reference number: JC2571; MUID:95011642
A:Accession: JC2571
A:Molecule type: DNA
A:Residues: 1-382 <PER>
A:Cross-references: EMBL:X79953; NID:g393391; PIDN:CA52139.1; PID:g393392
A>Note: This cellulolytic strain was isolated from the gut of termites
A:Gene: egls
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation

```


QY 497 PSPSPSPSSPSPSPSPSSPSPSPSPSPSPSSPSPSPSSPSPSPSPSPSPSSPSPS 556
Db 507 ----- 506
OY 557 PSSSPSPSPSPSPSSPSPSPSPSPSGGLKVOYKKNNDSAPGADQINKGCLQLVNTGSS 616
Db 507 -----PSGTFEV-----KAMIRNDSAFPARS----- 528
OY 617 SYDLSTVTVRWVFPTRDG--GSSTLVYVNCDAAMAACGNIRASFGSVNPATPRADTYLQLSFT 675
Db 529 ---LNNAKRVRVPTTDGFAPADVTLTSANSE--CG--AOSGAGVSAGTGLG-YVELSCV 579
OY 676 GGTLLAAGST---GEIQNRV-NKSDMSNETENDYSYG--TNTFQDMTKRVTVYNGVLV 729
Db 580 CODIHGCQSQHRRREIOFLRGTGPAGWN---PANDPSYTGILQTALAKASAITYTLYDGSTLV 636
OY 730 WTEPESGTPSPDPSPSPSPSPGGDVTPPSVPITGLVTVGVSSGSLAWMASTD- NVG 768
Db 637 WKPEPTGITTT-----DTTPPTTGPFPVATGVTTVGASLSLMAASTDAGSG 660
OY 789 VAHVWYV----NGVLVGOPVTYSFEDTLAAGTAATYVAAADAAGNTSAPTVPATY 844
Db 681 VAGVEYLRVQGTTQTLVGTTHAAAYILRLDLPETAISYYVKADAVGNVASAAATYFTT 740
OY 845 TSPSPSPPTPGTGYVDCTPFPG-PNQCVCIVSYVGDEYRVQTNEMNSSAOQCLTINTATGAWT 903
Db 741 DT-----TGETEPPPTTGPFPVASAVTS-----TGATLAM 770
OY 904 VTSTAMPSSGTGAPARTYPIRYKGCHMGNCNTTKNVGMPIQISOIGSAVTSMTSTOVSSGAY 963
Db 771 PST-----GDPAV--SGYDLRVQGTGTTTVVAOTVPTVTLGSLTPSTAVTYARAK 820
OY 964 DVAVIDWTNSTP-----TTTGQPNCTEIMINLWSRGVQOPGSGPATGTVTGHMWNWQO 1020
Db 821 NVAQGVSAISAAPVTFTTAAPVDT-----VAP-----TVPGTPVA----- 855
OY 1021 QTSWKIISVYLTPGATSISNLDKAIFADAAARGSLNTSDYLDIVEAGEFIW-----QQG 1075
Db 856 -----SNVATTGATL-----TWASTDSCG 875
OY 1076 OCL-GSNSFSVSVTSGTSSPTSPSPPTPPSPPTP-----SPSTPSP-SP 1120
Db 876 SGLAGYEVLRAVSGTQTOLVASPTTAVALAGLTAPATAYSIVRAKDAGANGVASVSSPYTF 935
OY 1121 TSPSSSGVACATYVVVNSDMDCSGFATYTVNTGSRATNGWTVAMSPGNGQVTVNTWNT 1180
Db 936 TTLPTPTSPSCVTVVSTNS-MWNGFPGSKIKITNGTTPILT-WLGRAPSPSQOYTQGWMSA 993
OY 1181 ALTOGASVATATNLSYNNVIRPOQOSTTFPGNSYSTGNAP 1221
Db 994 TMSQGTGVTVATGLSMNATLQPQGSIDICFNQSHPETNINP 1034

RESULT 13
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C.Species: Caenorhabditis elegans
C.Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #extl_change 18-Feb-2000
C.Accession: T34434
R.Gelabel, C.: Gallung, S.
submitted to the EMBL Data Library, December 1996
A.Description: The sequence of C. elegans cosmid K06A9.
A.Reference number: 221525
A.Accession: T34434
A>Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-2232 <GRI>
A.Cross-references: EMBL:080846; PIDN:AAC70890.1; GPSDB:GN00028; CESP:K06A9.1a
A.Experimental source: strain Bristol N2; clone K06A9
C.Genetics:
A.Name: CESP:K06A9.1a
A.Map position: X
A.Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 204

[illegible]

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:22:40 ; Search time 38.44 Seconds
(without alignments)
1236.930 Million cell updates/sec

Title: US-09-917-384-1
Perfect score: 6525
Sequence: 1 MERTQOGRNCRVORGTTRM.....GFNGSYSGTNAAPTLSCTAS 1228

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1655.5	25.4	872	1 GUXA_CELFI	P50401 Cellulomona
2	744.5	11.4	1742	1 GUNA_CALSA	P22334 Caldocellum
3	671.5	10.3	1331	1 MAMB_CALSA	P22333 Caldocellum
4	641	9.8	438	1 GUX3_AGABI	P49075 agarius bl
5	581.5	8.9	1090	1 GUXB_CELFI	P50899 Cellulomona
6	580	8.9	1045	1 GUXB_CELFI	P26225 Cellulomona
7	561	8.6	462	1 GUXB_FUSOX	P46236 fusarium ox
8	548	8.4	471	1 GUX2_TYRRE	P07987 trichoderma
9	543	8.3	562	1 GUN1_ACICE	P54583 acidotherma
10	519.5	8.0	1039	1 GUNB_CALSA	P10474 c endogluc
11	505	7.7	747	1 GUXB_CELFI	P50400 Cellulomona
12	493.5	7.6	5179	1 MOC2_HUMAN	002817 homo sapien
13	488	7.5	1853	1 CIPA_CLOTM	006851 clostridium
14	472.5	7.2	772	1 CIPB_CLOTM	001866 clostridium
15	443	6.8	1367	1 AMYH_YEAST	P08640 saccharomyc
16	416.5	6.4	474	1 VTP3_TTVIV	P19275 thermoprote
17	411	6.3	279	1 Y091_NPVOP	O10341 oryza pseu
18	406.5	6.2	555	1 GPL_CHLRE	O9f096 chlamydomon
19	404	6.2	441	1 GUN2_THERU	P26222 thermomons
20	386	5.9	145	1 YCBA_PAEFLA	P29718 paenibacill
21	378	5.8	499	1 GUN2_BACSU	P10475 bacillus su
22	377.5	5.8	700	1 GUNA_PAEFLA	P29719 paenibacill
23	376	5.8	456	1 GUNA_MICBI	P26414 microbispor
24	375.5	5.8	499	1 GUN3_BACSU	P23449 bacillus su
25	370.5	5.7	499	1 GUN1_BACSU	P07983 bacillus su
26	369.5	5.7	1664	1 SLPL_CLOTM	O06852 clostridium
27	355	5.4	1140	1 YW96_YEAST	O04893 saccharomyc
28	353.5	5.4	504	1 GUNB_ERWCA	O05995 erythrina car
29	348.5	5.3	880	1 GUN4_THERU	P26221 thermomons
30	347.5	5.3	1229	1 N121_HUMAN	O9y723 homo sapien
31	346.5	5.3	1848	1 CIPA_CLOTM	P26258 clostridium
32	344	5.2	505	1 GUNV_ERWCA	O47096 erythrina car
33	339.5	5.2	484	1 GUX_CELFI	P07986 cellulomona

34	335	5.1	699	1 CH11_BACCI	P20533 bacillus ci
35	335	5.1	1970	1 RPB1_MOUSE	P08775 mus musculu
36	333	5.1	1970	1 RPB1_HUMAN	P24928 homo sapien
37	324.5	5.0	1802	1 HKR1_YEAST	P41809 saccharomyc
38	322.5	4.9	2090	1 N214_HUMAN	P35658 homo sapien
39	313.5	4.8	449	1 APG_BRNA	P40603 brassica na
40	313	4.8	268	1 NS20_MEDTR	P93329 medicago tr
41	313	4.8	1306	1 MSB2_YEAST	P32334 saccharomyc
42	312	4.8	1199	1 N121_RAT	P52591 rattus norv
43	310	4.8	986	1 GUN2_CLOSR	P23659 clostridium
44	310	4.8	1150	1 APMD_PIG	P12021 sus scrofa
45	309.5	4.7	914	1 GUX2_CLOSR	P50900 clostridium

ALIGNMENTS

RESULT	ID	STANDARD	PRT	872 AA.
1	GUXA_CELFI			
AC	P50401:			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Exoglucanase A precursor (EC 3.2.1.91) (Exocellulohydrolase A)			
DE	(1,4-beta-cellulohydrolase A) (CBP95).			
GN	CBHA.			
OS	Cellulomonas flm1.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.			
OX	NCBI_Taxid=1708;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 484;			
RX	MEDLINE=94344030; PubMed=8065260;			
RA	Meinke A., Gilkes N.R., Kwan E., Kilburn D.G., Warren R.A.J.,			
RA	Miller R.C., Jr.;			
RT	"Cellulohydrolase A (CbaA) from the cellulolytic bacterium			
RT	Cellulomonas flm1 is a beta-1,4-exocellulohydrolase analogous to			
RT	Trichoderma reesei CBH II.";			
RL	Mol. Microbiol. 12:413-422(1994).			
RN	[2]			
RP	SEQUENCE OF 41-58.			
RX	MEDLINE=93209933; PubMed=8458833;			
RA	Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C., Jr., Warren R.A.J.;			
RT	"Cellulose-binding polypeptides from Cellulomonas flm1: endoglucanase			
RT	D (Cend), a family A beta-1,4-glucanase.";			
RL	J. Bacteriol. 175:1910-1918(1993).			
CC	- FUNCTION: THIS ENZYME HYDROLYSES 1,4-BETA-D-GLUCOSIDIC LINKAGES OF			
CC	CELLULOSE. WEAK ACTIVITY AGAINST CARBOXYMETHYLCELLULOSE, BACTERIAL			
CC	MICROCRYSTALLINE CELLULOSE AND BARLEY BETA-GLUCAN. HAS ALSO WEAK			
CC	ENDOGLUCANASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS WITH INVERSION			
CC	OF ANOMERIC CONFIGURATION.			
CC	- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages			
CC	in cellulose and cellotriose, releasing cellobiose from the non-			
CC	reducing ends of the chains.			
CC	- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.			
CC	- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN			
CC	(CBD).			
CC	- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL			
CC	HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: L25809; AAC36898.1; -			
DR	HSSP: P07986; IEXG			
DR	InterPro: IPR001919; CBD_2.			


```

RL  Appl. Environ. Microbiol. 57:694-700(1991).
CC  -1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC  ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC  DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC  CELLULOSE.
CC  -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC  linkages in cellulose.
CC  -1- PMR: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
CC  SITE FOR PROTEOLYSIS.
CC  -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC  E (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC  -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
CC  L (FAMILY 48 OF GLYCOSYL HYDROLASES).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  DR  EMBL; L32742; AAA91086.1; -.
CC  DR  EMBL; M36063; AAA72860.1; -.
CC  DR  EMBL; L01257; -. NOT_ANNOTATED_CDS.
CC  DR  PIR: A43745; A43745.
CC  DR  HSSP; P26221; 1TF4.
CC  DR  InterPro; IPR001956; CBD_3.
CC  DR  InterPro; IPR000556; Glyco_hydro_48.
CC  DR  InterPro; IPR001701; Glyco_hydro_9.
CC  DR  Pfam; PF00942; CBD_3; 3.
CC  DR  Pfam; PF02011; Glyco_hydro_48; 1.
CC  DR  Pfam; PF00758; Glyco_hydro_9; 1.
CC  DR  PRINTS; PR00844; GLHYDRASE48.
CC  DR  ProDom; PD001947; CBD_3; 2.
CC  DR  ProDom; PD011903; Glyco_hydro_48; 1.
CC  DR  PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
CC  DR  PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
CC  Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
KW  SIGNAL
FT  CHAIN 1 23
FT  SIGNAL 24 1742
FT  CHAIN 24 642
FT  DOMAIN 643 700 CATALYTIC 1.
FT  DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).
FT  DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).
FT  DOMAIN 904 1060 CELLULOSE-BINDING (BY SIMILARITY).
FT  DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).
FT  DOMAIN 1113 1742 CATALYTIC 2.
FT  ACT_SITE 396 396 BY SIMILARITY.
FT  ACT_SITE 434 434 BY SIMILARITY.
FT  ACT_SITE 443 443 BY SIMILARITY.
FT  CONFLICT 1545 1545 T -> A (IN REF. 2).
SQ  SEQUENCE 1742 AA; 193696 MW; 3F0699A213EED07 CRC64;

```

Query Match	11.4%;	Score 744.5;	DB 1;	Length 1742;
Best Local Similarity	22.8%;	Pred. No. 1e-22;		
Matches 281;	Conservative 162;	Mismatches 364;	Indels 423;	Gaps 46

```

0Y 29 AGVLAGANSIASIYPLMOMPAIAAHV-----DNRYAGATFEVNY-- 71
Db 173 AVAVETASLASIAASTVLDKDRNPTNATLYLOHAKDLYEPAEYTKSDSGYLANGYNSWGC 232
0Y 72 -----MAOEVOSEANOFNATLAKMRHVSTYSAVMMDRIAAINGVNGGPGCLTYLDA 125
Db 233 FYDELSTNA-----AWML- YLATND-----STYLFK 256
0Y 126 ALSQOQGTPEVIEYVYDLPGRCCALASNGE--LPATAGLOTYETQYIDPLASTLSN 183
Db 237 AESTYVO-NMPKISGSNIIDYMAHACMDVDVHGALLLAKITDKDQFYK-QIIESHLDWYTT 314
0Y 184 P-----KYSSLRITYITIEPDLSENAVTNMSI-----QACATAV-----PYEAGIEY 225

```

[illegible]

RESULT 3
MANB CALSA

MANB_CALSA STANDARD; PRT: 1331 AA.
 AC P25333;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Beta-mannanase/endorucanase A precursor [Includes: Mannan endo-1,4-beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)].
 CN MANA.
 OS Caldoceum saccharolyticum (Caldoceum saccharolyticus).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Thermotogaobacter group; Caldoceum saccharolyticum.
 OX NCBI_TaxID=44001;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93119139; PubMed=1476429;
 RA Gibbs M.D., Saul D., Luthi E., Bergquist P.L.;
 RT "The beta-mannanase from 'Caldoceum saccharolyticum' is part of a multidomain enzyme."
 RL Appl. Environ. Microbiol. 58:3864-3867 (1992).
 RP SEQUENCE OF 1-346 FROM N.A.
 RX MEDLINE=91247819; PubMed=2039230;
 RA Luehthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
 RT "Cloning, sequence analysis, and expression in Escherichia coli of a gene coding for a beta-mannanase from the extremely thermophilic bacterium 'Caldoceum saccharolyticum'."
 RL Appl. Microbiol. 57:694-700 (1991).
 CC -1- FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH MANNANASE AND ENDOLUCANASE ACTIVITIES.
 CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans, and galactoglucanmannans.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.
 CC -1- MICELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES CELSIUS.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY J (FAMILY 44 OF GLYCOSYL HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: L01257; AAA71887.1; -
 DR EMBL: M36083; AAA72861.1; -
 DR PIR: B43745; B43745.
 DR PIR: A48954; A48954.
 DR HSSP: 006851; INBC.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR001547; Glyco_hydro_F5.
 DR Pfam: PF00942; CBD_3; 2.
 DR Pfam: PF00150; cellulase; 1.
 DR ProDom: PD001947; CBD_3; 2.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR HydroLase: Glycosidase; Cellulose degradation; Signal;
 KW Multifunctional enzyme.
 FT SIGNAL 1 41
 FT CHAIN 42 1331
 FT DOMAIN 42 325
 FT DOMAIN 326 361
 FT DOMAIN 362 518
 FT DOMAIN 519 564
 FT DOMAIN 565 720
 FT DOMAIN 721 780
 FT DOMAIN 781 1331
 BETA-MANNANASE/ENDOLUCANASE A.
 CATALYTIC (MANNANASE ACTIVITY).
 PRO/SER/THR-RICH (PT BOX).
 SUBSTRATE-BINDING (POTENTIAL).
 PRO/SER/THR-RICH (PT BOX).
 SUBSTRATE-BINDING (POTENTIAL).
 PRO/SER/THR-RICH (PT BOX).
 CATALYTIC (ENDOLUCANASE ACTIVITY).

FT ACT SITE 162 162 PROTON DONOR (BY SIMILARITY).
 FT ACT SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
 FT CONFLICT 338 338 T -> P (IN REF. 2).
 FT CONFLICT 340 346 TPPTPT -> ROHORO (IN REF. 2).
 SQ SEQUENCE 1331 AA; 146892 MW; FFBG51B88F080 CRC64;
 Query Match 10.3%; Score 671.5; DB 1; Length 1331;
 Best Local Similarity 23.5%; Pred. No. 5.9e-20;
 Matches 243; Conservative 145; Mismatches 305; Indels 341; Gaps 43;
 240 MDAHSGWGMPPN-----ASGY-----VOEVGV--LMSIGVNGI----- 274
 DB LDTALGIRSGMNSRYVLSNGYRTKIPASVANIISLSLGRKAIILEVHDTGG 126
 QY 275 -DEFTVNTAN---YTPLK-----EPMTATQOVGQPEANFYQWNPDIIDEADYAD 323
 DB 127 EDGAGSLAQAVYWEKISVLDGNEDEYII--NINNEYGNNNYQWVNDTNNALKA-- 182
 QY 324 LSRVLAAGFPSSIGMLIDTLRNG--WGGPNEPTGS-----TA 360
 DB 183 ---LRDAGFKHTI--WVDAPNMGODMSMTMRDASIMEADPLRLVFSIHVGYNTA 236
 QY 361 TDVTFVNSKIDLRQHRGLMGNQAGLGQPPQASPT--DRPNHLDAYVWIKRPGES 417
 DB 237 SKVEEYI-KSFVD---KGL-----PLYGEGGHQHTDG-----DPDEE 270
 QY 418 DGTSAADPTTKKSDPMC-DPTVTTSYGLVNLALNSPIA-GQWPAQFDOLVANARPA 475
 DB 271 AIYRAVKQYKIGLFMSWCGNSSYGVLYDMVNMWDNNPTPMQGWKTN-----A 320
 QY 476 VPTSTSSPPPPPSASPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPPSSPSPP 535
 DB 321 IGT-----SPTPTPS 331
 QY 536 SSSP 594
 DB 332 TVPTPTPTPTPTPTPTATP-----TPPTPVSTPATSGIKVLYAN 373
 QY 595 NDSAPDNDQIKPGLQVNTGSSSVLDISTVTRYVETRDGSSSTLVYNCDMAAGCNIRA 654
 DB 374 KETNSTNTIRPLKLVNCGSSSIDLSRVTIRVYVDEGRASALS-DWAQGASNVTF 432
 QY 655 SFSVNPAPPTADTYQLSTFGT--LAAGSGEIQNVKNSKDSNFETNDYSI-GIN 711
 DB 433 KFKLSSVSGADYLYLEIFKSGAGQLQPKDGEIQMRNKMDSNMYNQNDMSWIOQS 492
 QY 712 TTFQDWTKYTVYVNGVLTGTEPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 771
 DB 493 TSGENKXTATYIDGVLVWQGEPSGATPAPATPTPTPT--VTP----- 536
 QY 772 SGSSVSLANNASTDNVGVAVHYNVRNGVLVQGPVTSTFDYGLAAGTAYTVVAADA 831
 DB 537 -----FTVTV----- 541
 QY 832 NTSAPSTPYTATTTSPSPSPPTGTVTDCITPBNQNGTVSVQGEYRV-----QTENKN 886
 DB 542 ---PPTVYAT--PPTPTPTPTPS-----TPATGGQIKVLYANKETNSTYT 583
 QY 887 SSAQOCL-TINTATGAMVTSTANFSGGTGAPATYSIYKGGCMWCTKKNQCMPIQISO 945
 DB 584 NTRPPLKLVNCGSSSIDLSKVTIR-----YW-----YTDGEH 617
 QY 946 IGSATVSWSTQVSGAYDAVDIMNTSPPTTGGQNGTEIMTWSRGVQVPPGSGQATP 1005
 DB 618 AGSAISDMA-----QIGASNVTFK-FVKLSSVSGADYLYEIG-FKSGAGQLDQGRKTGEI 671
 QY 1006 GYVYAAHTV-NWQSGQTSKTIISYLTPEGATISINDLKAIPADAAAGSLNTSDYILD 1064
 DB 672 QIRFNKSDMSNMYNQNDMSW-----IQSMTSYGENKXTATYID----- 709
 QY 1065 VEAQFEIWMGGQGLGNSFSVSYSVTSGTSSPTSPSPPTPTPTPTPTPTPTPTPTSSP 1124
 DB 1065 VEAQFEIWMGGQGLGNSFSVSYSVTSGTSSPTSPSPPTPTPTPTPTPTPTPTPTSSP 1124

DR PROSITE, PS00065: GLYCOSYL_HYDROL_F6_2: 1.
 KM Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 438
 FT DOMAIN 21 59
 FT DOMAIN 60 87
 FT DOMAIN 88 438
 FT ACT_SITE 215 215
 FT ACT_SITE 393 393
 FT DISULFID 28 45
 FT DISULFID 39 55
 FT DISULFID 170 229
 FT DISULFID 360 407
 FT VARIANT 133 133
 FT VARIANT 152 152
 FT VARIANT 244 244
 FT VARIANT 248 248
 FT VARIANT 398 398
 SO SEQUENCE 438 AA; 46209 MM; 002C973544893794 CRC64;

Query Match 9.8%; Score 641; DB 1; Length 438;
 Best local similarity 35.0%; Pred. No. 3.2e-19;
 Matches 162; Conservative 64; Mismatches 145; Indels 92; Gaps 14;

14 QRGTTMPAISKRLRAGVAGVASTAIVELAMQHAIATHTVNDNYPAGATFVNPYWA 73
 68 QPGTT-PRATTSGGTGTGTCGA-----GNYTGTWLSPFYA 104
 74 QEVOSEAAQNTNATLAAKMRVYSTYRAVMMDRIALINGVNGCGLTYYIDALSSQOQT 133
 105 DEVAQAAADISNPSLATKAASVAKIPFEVWFDTYAKV-----PDLGYLADARSKND-- 156
 134 TPPEYETIYDLPKDCALASNGELPATATAGLGTQYETQYIDPASTLSNPKYSLRIYT 193
 157 ---LVQIVYDLPKDCALASNGEFLSANDGLNKKY-NTYDQIAAQTK-QPDDSVYA 210
 194 IIEPDSLPNATNNNSIQACATAVPYEEGEGEYALTFLKHALPNVYIYMDAHSGLMGPNN 253
 211 VIEPDSLANLTNLTNVOGCANAQSAKYGVIYAVQKLNAY-GYMYIDAGHAGLWGPAN 269
 254 ASGYVOEKVAKLNASIGVNGIDGFEVYNTANTPLKEPMTATQVGCQPVESANFYQMP 313
 270 LSPAAQLEAFQIRBAGSPBRNLRGATINAVANFNALRAS-----SPDPIT-----OGNS 316
 314 DIDEADVAVDLSRLVAAGPSSIGMLIDLRLNMGNGPNEPTGPTATDVNTFVNQSKID 373
 317 NYDEHYTEALAPMLSNNGEPAH--FYVDQGRSG-----VON 351
 374 LROHRLGLNONGAGLQGPQASPTDPEFNALIDVYVMIKPGESDGTSAASDPTTGKSSD 433
 352 IRDQGMDCNKKAGAGFGGRP---FTNGSSSLIDAIYVKKPGEGCDGSDNSNP-----RFD 404
 434 PMSDPTTYTTSYGVLTNALPLNSPIAGQWEPAPFDOLVANARPAV 476
 405 SHCS-----LSDAHQAPAEAGTWQAVFETLIVANANPAL 438

RESULT 5
 ID GUXB_CELFI STANDARD; PRT; 1090 AA.
 AC P50899;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Exoglucanase B precursor (EC 3.2.1.91) (Exocellulobiohydrolase B)
 DE (1,4-beta-cellulobiohydrolase B) (CBP120).
 GN CBHB OR CENE.
 OS Cellulomonas flm1.
 OC Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococciaceae; Cellulomonadaceae; Cellulomonas.
 NX NCBI_TaxID=1708;
 GN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.

RC STRAIN-ATCC 484;

RX MEDLINE=96003896; PubMed=7575482;

RA Shen H., Giles N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;

RT "Cellulohydrolase B, a second exo-cellulohydrolase from the R

RT cellulolytic bacterium Cellulomonas fimi.;"

RL Biochem. J. 311:67-74(1995).

[2]

RP SEQUENCE OF 54-75.

RX MEDLINE=93209933; PubMed=8458833;

RA Melnke A., Giles N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;

RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase

D (Cend), a family A beta-1,4-glucanase.;"

RL J. Bacteriol. 175:1910-1918(1993).

[3]

RP SEQUENCE OF 54-78.

RX MEDLINE=94197708; PubMed=8147863;

RA Shen H., Tomme P., Melnke A., Giles N.R., Kilburn D.G.,

Warren R.A.J., Miller R.C. Jr.;

RT "Stereochemical course of hydrolysis catalysed by Cellulomonas fimi

Cend, a member of a new family of beta-1,4-glucanases.;"

RL Biochem. Res. Commun. 199:1223-1228(1994).

-1- FUNCTION: HYDROLYSE CELLULOSE TO A MIXTURE OF CELLOTRIOSE,

CELLOTRIOSE AND CELLOBIOSE, WITH ONLY A TRACE OF GLUCOSE. IT

CELLOTRIOSE TO CELLOBIOSE, BUT IT DID NOT HYDROLYSE CELLOTRIOSE.

HYDROLYSED CELLOPENTAOSE TO CELLOTRIOSE AND CELLOBIOSE. AND

HAS ALSO WEAK ENDOTRISACCHARIDASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS

WITH INVERSION OF ANOMERIC CONFIGURATION.

-1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages

in cellulose and cellobiose, releasing cellobiose from the non-

-1- reducing ends of the chains.

-1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.

-1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN

(CBD).

-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY 1 (FAMILY 48 OF GLYCOSYL

HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation-

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@sib-sib.ch).

DR EMBL; L38827; AAB00822.1; .

DR HSSP; P07986; IEXG.

DR InterPro: IPR001919; CBD_2.

DR InterPro: IPR003961; FN.III.

DR InterPro: IPR003962; FN.III.repeat.

DR InterPro: IPR000556; Glyco_hydro_48.

DR Pfam; PF00553; CBD_2; 1.

DR Pfam; PF00041; fn3; 3.

DR Pfam; PF02011; Glyco_hydro_48; 1.

DR PRINTS; PRO0014; FNTYPE.III.

DR PRINTS; PRO0844; GLHYDRLASE48.

DR PRODOM; PD011903; Glyco_hydro_48; 1.

DR SMART; SM00060; FN3; 3.

DR PROSITE; PS00561; CBD_BACTERIAL; 1.

KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.

FT SIGNAL 1 33

FT PROPEP 34 53

FT CHAIN 54 1090

FT DOMAIN 54 699

FT DOMAIN 700 785

FT DOMAIN 794 884

FT DOMAIN 891 978

FT DOMAIN 989 1090

FT ACT_SITE 513 513

FT DISULFID 990 1089

FT SEQUENCE 1090 AA; 114829 MW; 046BB9D956F2F399 CRC64;

Query Match 8.9%; Score 581.5; DB 1; Length 1090;

Best Local Similarity 21.8%; Pred. No. 1.7e-16;

Matches 306; Conservative 161; Mismatches 427; Indels 509; Gaps 63;

20 MPASIKRLRAGLAGAVSTAIVPLAMQHPRALTAHVDNPAGATFFNPVMAQEVQSE 79

1 MSSTTRRSAMWAATVGV--SSFLAVAGITPAIAAGACGP--AT-----VIVP 46

80 AANQTNATLAAK--MRVYSTSTAVWMDRIAAINGVGGPGLTTLDAALSOOGGTTPPEV 137

47 AASPRAAADVGEYQREFLAQY-----DKIK--DPANG-----YFSA-----QGIPYHA 87

138 IETIVYIDLP--GRDCAALASN-----GELPATAQL-QTET-----QYIIDPA 178

88 YETLWEAPADYEHETTSSEYSYMLWLEALYGVTDMDMPLNHAHMDTMEKMTPOSVDPT 147

179 SLTSPKYSRLRIVTIIERDSLPNAVTVN-----MSIQACATAVPYEGDIEVALTFL 230

148 NSFYPNPSFATYAPPEPNHPSSTPSQINSIGCTPDIGALKAT--YGADVQOMWL 203

231 HAIPVYIYMDAAHSG-WLGMPPNNAAGYQVQYKVLNLSI-----GVNGI 274

204 ADVNDIYRGATPGAGCTLGPATGTSFINTFQRGQESVWETVPQPSCEERKYGKNGY 263

275 DGFVNTANTATYTLKEPMTATQVGGQPYES--ANFTQWNPDIIDA----- 318

264 IDLEFTKDAISA--KWKYTSASDADARAVEAYVAN--QMAVEQCKADVAATVAKAKM 319

319 -DY-----AVDLRSRY-----AAGFPESIGMLDITLNGGNGNE----- 353

320 GYLKLTFLDKFKKIGCTSPICCAAGQGEAAHYLSYWMANGCATDSSGAMRIGSSH 379

354 -----PTGSTATDVTTFVNSKIDLRQHRGIMCNGAGL--- 389

380 AHFGYQNPPLAAMALSTDPLKTFKSPAKADMAAM--QRQLEFY---TWLQASNGGINGG 434

390 -----GQPPQASPTDFPNAHIDAYWIKPPGES--DGTPA-----ASDPT 428

435 ATNSMDGAVAGPAPACTPFYGGMGTEAPYVDPSPNRWFGMGQAVQVRAELYYASGNAQ 494

429 GKSPDPCDPYTTSGYVLTNALPNSPIAGQWFAQFDLVANAPVATSTSSSPPPP 488

495 AKK-----ILDKVVP-----WVAN-----ISTDASKVQVSE 522

489 PSPSASPS 548

523 LKWTGKPDTPWMAAATGNP----- 541

549 PSSSPS 608

542 -----GLTYE----- 546

609 QLVNTGSSSVDLSTVTVR---YMFTRDGGST---LYNCDWAMAGCGNIRASRGSVN 660

547 -VTSYGQDVGAADTARALLFYAAKSGTARDAKALLDIWA-----NQDPLGV- 597

661 PATPADYVLSFT-----GGTLAAGSGTGEION--RVKSDMSNFTETNDYSYGTNTTF 714

598 SAVETRGDKYRRDDYVANGDGIYIPSGWTGTMPNDVIKPGVSFLDINSFKKQPN-- 654

715 QWTKVTVTVVNVGLVWGTTP-----SGTSPSPSPSPSPSP 749

655 --WSKVQTFIDG---GAEPQRFYRHFMAQTVAGALADYARLFDDGTT--TP----- 699

750 SPSPGADVTPPVPGLVYTVGVSGSSVSLAAMASTDQNVGAHYNYRNGVLVGQPTVTSF 809

700 -----DTTAPVPPGLAGAVTSTREATISWASIDDTVTGYVDYRGAATKGTATTTTSF 753

810 TDTGLAAGTAYTVAAVDAAGTASPTVATTTSPSPSPPTPTGTCTDPTCPGNONG 869

754 TDTGLTASTAYTVRAFDAGNVASPSAALTVTTRKA-----TPSDTT-----APSVA 802

870 VTSVQGDERTVQTNEMNSSAQOCLTINTATGAMTVSTANFSGCTGGAATPYSTIKGCHW 929

QY 314 DIEDADYADVLSRLVAGPSSIGMLIDPLRMGSGPNEPTGSTATDVNTFVNOSKID 373
DB 338 NYDEORY-INAAPVLAEOGMSNVKFIYDGRSG-----KQPTG----- 375
QY 374 LROHRCIMQNGAGLGQPPQASPTDFPNAHLAVYVIRKPGESDGTSAASDPPTGKSD 433
DB 376 -QAAGQGMCAKGTGELRSTNGD---ALADFEVWVKGGESDGTG---DTSARYD 427
QY 434 PMCDPTTTSYGVLTNALPNSPIAGWFPAPQFDOLVANAPA 475
DB 428 YHCG-----LDDALRKAPRAGTWFOAYFEQLDANAPS 460
RESULT 8
GUX2_TRIRE STANDARD: PRT: 471 AA.
AC P07987;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Exoglucanase II precursor (EC 3.2.1.91) (Exocellobiohydrolase II)
GN CBH2
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_Taxid=51453;
RN SEQUENCE FROM N.A.
RC STRAIN-VTT-D-80133;
RA MEDLINE-87248061; PubMed-3596237;
RT Teeri T.T., Lehtovaara P., Kauppinen S., Salovuori I., Knowles J.;
RT "Homologous domains in Trichoderma reesei cellulolytic enzymes: gene
RL sequence and expression of cellobiohydrolase II."; Gene 51:43-52(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-QM9414 / RUT C-30;
RA Chen C.M., Gritzall M., Stafford D.W.;
RT "Nucleotide sequence and deduced primary structure of
RL cellobiohydrolase II from Trichoderma reesei."; Biotechnology 5:274-278(1987).
RN [3]
RP SEQUENCE OF 25-44.
RA Faegerstaem L.G., Pettersson L.G.;
RT "The 1,4-beta-glucan cellobiohydrolases of Trichoderma reesei QM
RL 9414."; FEBS Lett. 119:97-100(1980).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE-90333255; PubMed-2377893;
RA Rouvinen J., Bergfors T., Teeri T.T., Knowles J.K.C., Jones T.A.;
RT "Three-dimensional structure of cellobiohydrolase II from Trichoderma
RL reesei."; Science 249:380-386(1990).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE-97029636; PubMed-8875646;
RA Koliyala A., Reinikainen T., Ruohonen L., Valkeajarvi A.,
RA Claeyssens M., Teeman O., Kleywegt G.J., Szardenings M., Rouvinen J.,
RA Jones T.A., Teeri T.T.;
RT "The active site of Trichoderma reesei cellobiohydrolase II: the role
RL of tyrosine 169."; Protein Eng. 9:691-699(1996).
CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) EXOCELLULOBIODHROLASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) ENDOCELLULOBIODHROLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-

CC reducing ends of the chains.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: T. REESEI PRODUCES TWO DIFFERENT
CC EXOCELLULOBIODHROLASES. THEY ARE UNIQUE IN THAT THEY CAN HYDROLYSE
CC CRYSTALLINE CELLULOSE IN THE ABSENCE OF ENDOCELLULOBIODHROLASES.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M16190; AAA34210.1; -;
DR EMBL: M55080; AAA72922.1; -;
DR PIR: A26472; A26472.
DR PIR: A26160; A26160.
DR PDB: 3CBH; 15-JAN-91.
DR PDB: 1CB2; 19-MAR-99.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR001524; Glyco_hydro_6.
DR Pfam: PF00734; CBD_1; 1.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PR00733; GLYHYDRLASE6.
DR ProDom: PD001821; CBD_fungal; 1.
DR ProDom: PD003733; Glyco_hydro_6; 1.
DR SMART: SM00236; fCBD; 1.
DR PROSITE: PS00562; CBD_FUNGAL; 1.
DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE: PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KM Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal;
KM 3D-structure.
FT SIGNAL 1 24
FT CHAIN 25 471 EXOGLUCANASE II.
FT DOMAIN 25 65 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 66 106 LINKER.
FT DOMAIN 107 471 CATALYTIC.
FT ACT_SITE 199 199 PROTON DONOR.
FT ACT_SITE 245 245 NUCLEOPHILE.
FT DISULFID 34 51 BY SIMILARITY.
FT DISULFID 45 61 BY SIMILARITY.
FT DISULFID 200 259
FT DISULFID 392 439
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 111 111 O-LINKED (MAN. . .).
FT CARBOHYD 121 121 O-LINKED (MAN. . .).
FT CARBOHYD 130 130 O-LINKED (MAN. . .).
FT CARBOHYD 133 133 O-LINKED (MAN. . .).
FT CARBOHYD 134 134 O-LINKED (MAN. . .).
FT CARBOHYD 139 139 O-LINKED (MAN. . .).
FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .).
FT CARBOHYD 334 334 N-LINKED (GLCNAC. . .).
FT MUTAGEN 199 199 D->A: 208 OF WILD-TYPE ACTIVITY.
FT MUTAGEN 245 245 D->A: NO MEASURABLE ACTIVITY.
FT CONFLICT 359 359 P -> R (IN REF. 2).
FT CONFLICT 449 449 P -> A (IN REF. 2).
SQ SEQUENCE 471 AA; 49653 MW; C4711BC35B1BD88 CRC64;
Query Match 8.48; Score 548; DB 1; Length 471;
Best Local Similarity 31.28; Pred. No. 1; De-15;
Matches 156; Conservative 60; Mismatches 176; Indels 108; Gaps 17;
QY 7 SGNRCRYR-----GTRMPAISRLRAGVLAAGVIASTIAP-----LAMQ 48
DB 47 SGRICVSNDRYSQCLGAASSSSSTRAASTTS-----VSPITSSSSATPPPGSTTTR 101
QY 49 HPAL--AATHVNDPVAGATFEFVNPVWAOEVSANQTNATLAAKRVVSVYSTAVAMD 105

```

DB 102 VPPVSGSRTATYSGNPFVCTPMAWNAVYASVSSLAIPSLTGM:RAAAAVAPSPFMD 161
OY 106 -----RIAINGVNGSGFGLTYYDADALSQOQGTTPVEIYVYDLPGRCALAS 155
DB 162 TLDKTPLEQOTLADIRTKANKNG--NYAG-----QYVYVDLPDRCAALAS 205
OY 156 NGELPATAAGLOTFYOTYVITPIASILSNPKYSRLIVTIEPDSLPNATNMSIQACATA 215
DB 206 NGEYSIADGGVAKY-NVIDTRQIV--VEYSDIRTLVIEPDSLANLTNGTPRCANA 262
OY 216 VPPYEGIEVALTKLHAIPNVYIYMDAAHSGWLGMPNNAAGVQYQKLNASIGVNGID 275
DB 263 QSAYLECINAYATQLN-LPNVAMYLDAGHAGLGWPANDPAQLPANYKNAASPRALR 321
OY 276 GCVYTNANTPLKEPPMTATQOVGPVESANFYQWNPDIADADYAVDLYSLVAGFPS 335
DB 322 GIATVNAVYNN-----GWNITSPSPSYTOGNNAVYNEKLYIHAIGLPLANGW-S 367
OY 336 SIGMLIDTLRNGMGPNPEPTSTATDVNTFVNOSKIDLRQHRGIMCNGNGAGIQPQA 395
DB 368 NAFETIDGGRSG---KQPTG-----QQQWGMWCNVYIGTGFGRPSA 405
OY 396 SPTDEPNALHDAVYWIKPPGESDGTSAASDPPTGKKSDPMCDPTWTSYGLTNALPNSP 455
DB 406 NTGD---SLDGSFVWVKGGECDDTSDSAP---RFDSHC-----ALPDALQRPAP 449
OY 456 IAGNFPAPFDOLVANARPA 475
DB 450 QAGAMFOAYFVQLTNNAPS 469

RESULT 9
GUNL_ACICE STANDARD: PRT: 562 AA.
AC P54583;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)
DE (Cellulase E1) (Endocellulase E1).
OS Acidothermus cellulolyticus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Frankineae; Acidothermaceae; Acidothermus.
OX NCBI_TaxID=28049;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 43068 / 11B;
RA Layton R.A., Himmel M.E., Thomas S.R.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RX X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.
RX MEDLINE=96346058; PubMed=8718854;
RA Sakon J., Adey W.S., Himmel M.E., Thomas S.R., Kaplun P.A.;
RT "Crystal structure of thermostable family 5 endoglucanase E1 from
RT Acidothermus cellulolyticus in complex with cellobiose."
RT Biochemistry 35:10648-10660(1996).
CC -1- FUNCTION: THERMOSTABLE ENZYME WITH AN OPTIMAL TEMPERATURE OF 81
CC DEGREES CELSIUS. HAS A VERY HIGH SPECIFIC ACTIVITY ON
CC CARBOXYMETHYLCELLULOSE.
CC -1- CATALYTIC ACTIVITY: Endohydrolase of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL contribution -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: U33212; AAA75477.1; -
DR PDB: 1ECE; 14-OCT-96; CBD_2.
DR InterPro: IPR001919; CBD_2.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF00553; CBD_2; 1.
DR Pfam: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5_1.
KM Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.
FT SIGNAL 1 41
FT CHAIN 42 562
FT DOMAIN 42 400 ENDOGLUCANASE E1.
FT DOMAIN 401 461 CATALYTIC.
FT DOMAIN 462 562 PRO/SER/THR-RICH (LINKER).
FT ACT_SITE 203 203 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 323 323 PROTON DONOR.
FT DISULFID 75 161 NUCLEOPHILE.
FT DISULFID 209 212
SQ SEQUENCE 562 AA; 60747 MW; 84E6256406A35041 CRC64;

Query Match 8.3%; Score 543; DB 1; Length 562;
Best Local Similarity 35.4%; Pred. No. 3e-15;
Matches 138; Conservative 48; Mismatches 104; Indels 100; Gaps 16;

OY 869 GYTSVQGDYRYRQTNEM-----NSSAQOCLTINTAT-----GAWYVSTANFSGCTGAPAT 919
DB 243 GVQSYNGDSY-----WMGNIQAGQYFVNLVNPRLVYSADYATSYTP-QTWFSQPT 295
OY 920 YPSIYKGC---HWGNCCTKRVNGMPIQISQISAVTSWSTTVSSGAYDAYDITNSTPT 976
DB 296 FPNMNPGLIMKRWGLFNFNIA-PWILGFEGLTDS----- 330
OY 977 TTGQPNGETEIMTILNSRGVQPGFSOTATGTVAGHTMNMWQGOQTSWKIISYVLPQAT 1036
DB 331 TYDQ-----TWLKITL-VQYLRTAQYAD--SFQWTFW-----SWN----- 363
OY 1037 SISNDLKAIFADAAARGSLNSTD-----YLDVEAFETIMQGGIGLSNFSFVS 1086
DB 364 -----PDSGDTGILKMDQYVDTYKDGIALPKSSIFDPVGSASPSQSPS 412
OY 1087 VT-----SGTSSPTSPSPPTPPSPPTP--TPSPSPPTSPSPSP--SSSGVACRTATVYN 1138
DB 413 VSPSPSPSSASRTPTPTPTPTASPTPTLTPTATPTPTPTASPTPTASGACRTASTOVN 472
OY 1139 SDWSGFTATVTVTNTGSRATNGVTWAMSEFGNQTVTNTWNTALTOGASATATNLSTNN 1198
DB 473 SDWNGFTVTVAVTNSGSVATKTWTSWTFEGNQTTNSMAAATVQNGSVYARMSYNN 532
OY 1199 VTPOGQSTFFGNGSVSGTNAAPTLSCTAS 1228
DB 533 VTPOGNTTFGFASTYGSNAAPTVAACAAS 562

RESULT 10
GUNB_CALSA STANDARD: PRT: 1039 AA.
AC P10474;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endoglucanase/exoglucanase B precursor (includes: Endoglucanase
DE (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)
DE (Cellulohydrolase); Exoglucanase (EC 3.2.1.91) (Exocellulohydrolase)
DE (1,4-beta-cellobiohydrolase)).
DE CELB.
OS Caldicellum saccharolyticum (Caldicellulosigrup saccharolyticus).
OC Bacteria; Firmicutes; Bacilli; Clostridium group;
OC Thermoclostracter group; Caldicellulosigrup.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098398; PubMed=2789517;
RA Saul D.J., Williams L.C., Love D.R., Chamley I.W., Bergquist P.I.;

```

```

RT *Nucleotide sequence of a gene from Caldocellum saccharolyticum
RT encoding for exocellulase and endocellulase activity.*
RL Nucleic Acids Res. 17:439-439(1989).
CC -1- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL
CC DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS
CC AN ENDOGLUCANASE.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellobiose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO CELLULASE FAMILY
CC F (FAMILY 10 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO CELLULASE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X13602; CAA31936.1; -.
DR PIR: S02711; S02711.
DR HSSP: Q06851; INBC.
DR InterPro: IPR001956; Glyco_hydro_3.
DR InterPro: IPR001000; Glyco_hydro_10.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS: PR00134; GLHYDRLASE10.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE: PS00599; GLYCOSYL_HYDROL_F5; 1.
DR Cellulose degradation; Hydrolyase; Glycosidase; Repeat;
KW Multifunctional enzyme; Signal.
FT SIGNAL 1 28
FT CHAIN 29 1039
FT DOMAIN 376 416 THR/PRO-RICH, TANDER REPEATS OF T-P.
FT DOMAIN 417 570 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 571 618 THR/PRO-RICH, TANDER REPEATS OF T-P.
FT ACT_SITE 177 177 PROTON DONOR (POTENTIAL).
FT ACT_SITE 285 285 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 792 792 BY SIMILARITY.
SQ SEQUENCE 1039 AA; 117641 MW; 0E0378171594DAB CRC64;

```

```

Query Match 8.0%; Score 519.5; DB 1; Length 1039;
Best local Similarity 23.6%; Pred. No. 4.6e-14;
Matches 133; Conservative 98; Mismatches 179; Indels 219; Gaps 21;

```

```

OY 832 NTSAPSPVATATTTSPSPPTGTVDCTPGFNGVTSVQGEYVQTNEMSSAQ 891
OY 597 -----PTAPQGTATPPTLPNPPPTSSIPDD-----INDMLYVSGN 635
OY 892 CLTINTATGAMVYSTANFSGCTGAPATYPSIYKCGHWGCTTKN-----VGM 939
OY 636 KIVDKDGRPWLGINNFVNTG-----TNVEFGV-W-SCNLKDTLAETIANRGFNLLRV 687
OY 940 PIDISGASVTSMTSTQVSSGAY---DVAYDITKNTSPPTTGPN----- 982
OY 688 PIS---AELILNM-----SGGTPKPNINYY---NPELEGKNSLVEFDIVYQCKEV 734
OY 983 GTEIMILNLRSGVQVPGSGATGATVAGHTWVWQOQTS----- 1023
OY 735 GLKIMLDIHS-----IKTDMGHILPYWDEKPTPEDFYKACEMITRYKND 782
OY 1024 -----WKIISYVLTPGATVSTSLDLKAIKPDAAARGLNTSDVLLDVAG 1068
OY 783 TIIAFDLKNEPHGKPMQDTTFARKWNSTDINNMKYA---AETCAKRLININPILLIIEG 839
OY 1069 FE-----IMGGGGLGSNSFSVSTGTSPTPSP 1098
OY 840 IEAVPKDDVTWTSKSSSDYSTWVGMLRGVRRKPIPLKRYONKRVYSP 888

RESULT 11
ID GUND_CELFI STANDARD; PRT: 747 AA.
AC P50400;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
GN CEND.
OS Cellulomonas fimi.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococccineae; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93209933; PubMed=8458833;
RA Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C., Jr., Warren R.A.;
RT "Cellulose-binding polysaccharides from Cellulomonas fimi: endoglucanase
RT D (Cend), a family A beta-1,4-glucanase."
RU J. Bacteriol. 175:1910-1918(1993).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PATHWAY: CELLULOSE DEGRADATION.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L02544; AAA23089.1; -.
DR HSSP: P07986; IEXG.
DR InterPro: IPR001919; CBD_2.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003962; FnIII_repeat.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF00553; CBD_2; 1.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00041; fn3; 2.

```


DR	PROSITE; PS01185; CTCK_1; 1.
DR	PROSITE; PS01225; CTCK_2; 1.
KW	Glycoprotein; Repeat; Signal.
FT	SIGNAL 1 20 POTENTIAL.
FT	CHAIN 21 5179 MUCIN 2.
FT	DOMAIN 1401 1747 APPROXIMATE REPEATS.
FT	REPEAT 1401 1416 1.
FT	REPEAT 1417 1432 2.
FT	REPEAT 1433 1448 3.
FT	REPEAT 1449 1464 4.
FT	REPEAT 1465 1471 5.
FT	REPEAT 1472 1478 6.
FT	REPEAT 1479 1494 7A.
FT	REPEAT 1495 1517 7B.
FT	REPEAT 1518 1533 8A.
FT	REPEAT 1534 1556 8B.
FT	REPEAT 1557 1572 9A.
FT	REPEAT 1573 1596 9B.
FT	REPEAT 1597 1612 10A.
FT	REPEAT 1613 1635 10B.
FT	REPEAT 1636 1651 11A.
FT	REPEAT 1652 1675 11B.
FT	REPEAT 1676 1683 12.
FT	REPEAT 1684 1699 13.
FT	REPEAT 1700 1715 14.
FT	REPEAT 1716 1731 15.
FT	REPEAT 1732 1747 16.
FT	DOMAIN 4815 4886 VMP_C 1.
FT	DOMAIN 4924 4991 VMP_C 2.
FT	DOMAIN 5075 5160 CTCK.
FT	DISELFD 5075 5122 BY SIMILARITY.
FT	DISELFD 5089 5136 BY SIMILARITY.
FT	DISELFD 5098 5152 BY SIMILARITY.
FT	DISELFD 5102 5154 BY SIMILARITY.
FT	DISELFD ? 5159 BY SIMILARITY.
FT	CARBOHD 163 163 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHD 670 670 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHD 770 770 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHD 894 894 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHD 1139 1139 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHD 1154 1154 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHD 1215 1215 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHD 1230 1230 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHD 1246 1246 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHD 1787 1787 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 1820 1820 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 4339 4339 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 4351 4351 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 4362 4362 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 4373 4373 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 4422 4422 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 4438 4438 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 4502 4502 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 4616 4616 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 4627 4627 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 4752 4752 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 4787 4787 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 4881 4881 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 4888 4888 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 4955 4955 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 4970 4970 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 5019 5019 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 5038 5038 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CAROHDH 5069 5069 N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CONFLICT 1351 1351 H -> L (IN REF. 3.)
FT	CONFLICT 1412 1412 T -> S (IN REF. 3.)
FT	CONFLICT 1449 1449 L -> P (IN REF. 3.)
FT	CONFLICT 1504 1504 M -> T (IN REF. 3.)
FT	CONFLICT 1492 1492 G -> S (IN REF. 2.)
SQ	SEQUENCE 5179 AA; 540295 MW; 85CDV57JTB9A55633 CRC64;

```

Query Match 7.6%: Score 493.5: DB 1: Length 5119;
Best Local Similarity 22.6%: Pred. No.2.3e-12;
Matches 219; Conservative 114; Mismatches 373; Indels 265; Gaps 33;

392 PQASPTDFPNAHLDAVYWKIPGGS---DGSAAADPTTGKSKSDPKMDPTYTSGVLT 448
Db PMTTP-----ITPPASTTLLPRTTSSPTTTTTPP-----PTTTPSPPTT 1545
QY 449 NALPNSPIAGOWFPAQFDOLVANARPAVSTSSSPP-PPSPGASPSPP-----SP 501
Db 1546 PTP-----PSTTTLPTPTTSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1596
QY 502 SPSSSP-----SPSSSPS-----PSPSPSSSPSSSPSS 538
Db 1597 PPTTTPSPPTTTPPTTTPPTTTPSPPTTPITPTTTLPTTTPSSPPTTTPPTTTP 1656
QY 539 PSP-----SPSP-----SPSSSPS-----PS 564
Db 1657 PSPPTTTPSPPTTTPPTTTPPTTTPSSPTTTPSPPTTTPPTTTPSSPTTTPSS 1716
QY 565 -----SPSPSSSPSSPPTSSPSPVSGGLKVQKKNDSAP-GONQIKPGLQLVNTHSSSD 619
Db 1717 STTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1776
QY 620 LST-----VTRYW-----FTDGGSGSLVYN-C-DMAAMCGNIRASFGSVNPA 663
Db 1777 PPTTTPCVPCLNNTGMLDSCKRPFHRRGQGTGLGVCGGGMA-----NISC 1831
QY 664 PRADTYLQISFGGTLAAGSGTGEIONRYNKSMDSNFTETNDYSYGTN----TFQDM 718
Db 1832 P-----IGQGGQVYVCDVSGLCKNKNDQKRGVYPMACFLANEINQCCCECV 1884
QY 719 KTVYVNGVLWGTPESSGSPPTSPSP-----SPSPSGGVTPPS 761
Db 1885 MTTITTENPTPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1944
QY 762 VPTGLVTVGSSSVSLAMNASTDNGVAHYNVRNGVLVGPITYSFTDGLAGT--- 818
Db 1945 PPTTTPPTTTPPTTTPPTTTPPTTTP-----TTTTPVTPPTTTPGQTP 1988
QY 819 -----AYTVAANDAGNTSAPSRPVNATTTSSPSPTPG-----TIVDCTFG 864
Db 1989 TPTTPTTTPVTPPTTTPGQTPPTTTPPTTTPPTTTPPTTTPGQTPPTTTPPT 2048
QY 865 PNONGVTSVOGDEYRVQTEEMNSAQQCLTITATGAMVSTANFSGCGGAPATVPSI 924
Db 2049 PPTTTP-----QPTTTPPTTTPVTPPTTTPGQTPPTTTPPTTTPVTPPTTP 2098
QY 925 KCHMGNCCTKVNQAPIQISQIGASVTSKTSQVSSGAYDAVDIWTNSTPTTGO 984
Db 2099 -GTGPTTTPPTTTPVTPPTTTPGQTPPTTTPPTTTP-----TTPVTPPTTTP 2152
QY 985 ELIMILNSRGVQPGSGSOTAGVTVAAGHWMNWQOQSMKLIISVLPGASISNLD 1044
Db 2153 PTTTTTTPPTTTPPTTTPGQTPPTTTPPTTTP-----TTPPTP----- 2189
QY 1045 AIFADAAAGSLNTSDYLLDVEAGEEIMOGGGLSNSPSVTSQTSAPTPSPPTTP 1104
Db 2190 -----TGQTPPTTTPPTTTPVTPPTTTPPTTTPPTTTPPTTTPPTTTP 2218
QY 1105 SPTP-----TPSPPTSPSPSTSSSSGACRAITYVNSDMGSGFAITYVNTGSR 1157
Db 2219 TPTTTPPTTTPVTPPTTTPGQTPPTTTPPTTTPVTPPTP-----TGQTPPTTTP 2270
QY 1158 ATNGTIVAMSFQNGQTVYVNTATLQSGASVATNLISNNYIQPGSTTFGNSYS 1217
Db 2271 TTTTTPPTTTPGQTPPTTTPPTTTPVTPPTTTPGQTPPTTTPPTTTPVTPPTT 2330
QY 1218 NAAPLISCTAS 1228
Db 2331 QPTTTPPTTTP 2341

```



```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SS288C / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentsles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.,
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE-87194600; PubMed-3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STA1."
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN-SP101-1C;
RX MEDLINE-89031230; PubMed-3141213;
RA Padro J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STA2 and SGA genes
RT from Saccharomyces cerevisiae."
RL FEBS Lett. 239:179-184(1988).
CC -1 CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1 SIMILARITY: TO S.POMBE SPC215.13.
CC -1 SIMILARITY: SOME. TO S.POMBE SPC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, Z38061; CA86176.1; -
DR EMBL, M16164; AA35014.1; -
DR EMBL, M16165; AA35015.1; -
DR EMBL, X13857; CA32069.1; -
DR PIR, B26877; B26877.
DR PIR, B26877; B26877.
DR PIR, S48478; S48478.
DR SGD, S0001458; MOC1.
DR Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KM Signal; Multigene family.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1367 GLUCOAMYLASE SI/S2.
FT DOMAIN 210 1367 SER/THR-RICH.
FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD1AA9D CRC64;

```

Query Match 6.8%; Score 443; DB 1; Length 1367;
 Best Local Similarity 23.1%; Pred.No.6.3e-11;
 Matches 243; Conservative 132; Mismatches 395; Indels 282; Gaps 38;

```

QY 351 PNEPTGPTATVNTFVNGSKIDLQHRGLMCNONGAGLGGPPQASPTDFPNAHDAIYV 410
DB 398 PSSSTTESSAPVTSSTESS-----SAPVTSSTESSA----- 432
QY 411 IRPPESDGTSAADP-----TTGKSDPCDPTT-----SYGLTNALNPSPIAGQWFA 463
DB 433 ---PTSSSTTESSAPVTSSTESSAPVPTPSSSTTESSAPVTSSTESSAPVPTPS 489
QY 464 QFDQVANAARPAVPTSTSSPPPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 513
DB 490 SSTSSAPVTSSTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTSSTES 549

```

```

QY 514 SSPPSPSPSPSPSPSS-----PSPSPSS-----PSPSPSPSPSPSS- 552
DB 550 SSAPVPTPSSSTTESSAPVPTVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSA 609
QY 553 PSPSPSS-----PSPSPSPSPSPSS--PSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 592
DB 610 PAPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTT----- 661
QY 593 KNDASAPGDNQIKPGIQLVNTGSSSVLSTVVRVFTRDGSSSTLVYNCDMAAGCGNI 652
DB 662 -ESSSAP-----VPTPSSSTTESSAPVTSSTTESSAPV----- 696
QY 653 RASFGVNPATPTADYIQLSTFGTLAAGSTGELQNKVNSDMNSNFTETDYSIGTWT 712
DB 697 SSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSS 756
QY 713 TFQDMTKVYVYNGVLWGTPE-----SGTSPSPSPSPSP-----PSPSPSPG 755
DB 757 TESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSS 816
QY 756 DVT---PSPVP-----TGLVVTGVSGS---SVSLAMNASIDNVGVAHYV--YRNGVL 800
DB 817 NITSSAPSTPSSSTTESSAPVPTPSSSTTESSAPVSSSTTESSAPVSSSTTESSAPVPTPSS 876
QY 801 VQOPVTSFTDGLAGTAYITVAADAGN---TSAPST-----PTATTSBSPSP 851
DB 877 SSAPSSIPSPSTTESSSTGTGTVTVPSSSKYPGOTETVSVSTETETIVPKRTTSTVPTPT 936
QY 852 TPTGTVTCTGTPGPNQNGVTSVQGDVRYQVTNEMNSAQOCLTINATGAMTVSTANFSG 911
DB 937 TITTTT---CSGTGNAGTIT--SGCSPKVTITVTPTTTTSTTSTT---TITTVCGST 990
QY 912 GTCGAPATYPSIYKCGCHWGCNCTTKNVGMPDIQISQISAVTSNSTQVSSGAYDAVDWT 971
DB 991 GINSAGET-----TSGCSPKRTTITVPCSTSPSEFASSTT----- 1026
QY 972 NSTPTTGTGPNQELIMIMWNSRGVQPPSSQATGCTVAGHWNWQSGQTSMKIISYVL 1031
DB 1027 -TSPPTVTVTVTVTVTVTVTSTRTKPGGEITTTT-----KNIPPTV-LTTIAP 1075
QY 1032 TPGATISIMLDKALFADAAANG-----SLNTSDY-----LL 1063
DB 1076 TSVTITVTNFTPTTITTTVCSGTNAGETSGCSPKVTITVPCSTGCEITTEATTV 1135
QY 1064 DVEAGFEIWOQGGGGSNSFSVSVTSSTSPSP-----SPSPPTPSPPTPSP-SPT 1115
DB 1136 TTAVTITVTVTSTSTGTNAGKTTGTGYSVPTVVTTLAPSAVPTPATNNAVPTTITTT 1195
QY 1116 PSPSPSPSSSGVACRATVYVNS-----DMSGRTATVTV-----TNTGS 1156
DB 1196 ECSAATNNAAGETTSVCSATVIVSSASAGENTAPSATVPTTATPTVTITTESSVGTNSAG 1255
QY 1157 RATNGWT-----VAMSGNGQVNTYNNWNTA-----LTQSGASYATNU 1194
DB 1256 ETTTGTITTSIPTTYITTLIPSGNKAKEVYATATNPISITITQSLATTAASASVAPV 1315
QY 1195 SYNNAVIOGQSTTFGNGSYSGTNAAPTLTCT 1226
DB 1316 TSPSLTGLPQSA-----SGSAVATVSVPSISST 1343

```

Search completed: August 29, 2002, 16:38:31
 Job time: 951 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:21:50 ; Search time 125.36 Seconds

(without alignments)
1694.622 Million cell updates/sec

Title: US-09-917-384-1

Perfect score: 6525
Sequence: 1 MERTQSGRCNRGRTGTRM.....GFNGSYSGTNAATLSTCTAS 1228

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seags, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Database : Listing first 45 summaries

- 1: SP archaea:*
- 2: SP bacteria:*
- 3: SP fungi:*
- 4: SP human:*
- 5: SP invertebrate:*
- 6: SP mammal:*
- 7: SP mhc:*
- 8: SP organelle:*
- 9: SP phage:*
- 10: SP plant:*
- 11: SP rodent:*
- 12: SP virus:*
- 13: SP vertebrate:*
- 14: SP unclassified:*
- 15: SP rivirus:*
- 16: SP bacteriophage:*
- 17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1358	20.8	683	16	09PDW2	09pdw2 xyella fas
2	1051	16.1	579	2	086730	086730 streptomyc
3	1015.5	15.6	454	2	09KH72	09kh72 thermomonos
4	1000	15.3	596	2	060029	060029 thermomonos
5	743.5	11.4	1711	2	P96311	P96311 anaerocellu
6	737.5	11.3	930	2	09RFX5	09rfx5 caldicellul
7	722.5	11.1	351	2	09S3V4	09s3v4 cellulomona
8	708.5	10.9	1751	2	09AGC4	09agc4 caldicellul
9	705	10.8	490	3	096V97	096v97 orpiniomyces
10	694	10.6	491	3	096V98	096v98 orpiniomyces
11	679	10.4	921	2	091818	091818 caldicellul
12	675	10.3	1779	2	052374	052374 caldicellul
13	669.5	10.3	381	2	09RKY3	09rky3 streptomyc
14	668.5	10.2	381	2	054331	054331 streptomyc
15	663.5	10.2	1426	2	09X3P6	09x3p6 caldicellul
16	654	10.0	377	2	008468	008468 streptomyc

17	653	10.0	1770	2	09X3P5	09x3p5 caldicellul
18	649.5	10.0	382	2	059663	059663 streptomyc
19	618.5	9.5	460	3	002321	002321 phanerocha
20	618	9.5	371	2	09KIH1	09kih1 streptomyc
21	597.5	9.2	444	3	096VU2	096vu2 lentinula e
22	590	9.0	1000	2	024820	024820 thermophil
23	589.5	9.0	443	3	09CIR4	09cir4 lentinula e
24	582	8.9	436	3	09P8N1	09p8n1 coriolinus ve
25	569	8.7	465	3	096RP4	096rp4 pleurotus s
26	565.5	8.7	476	3	09CIS9	09cis9 humicola in
27	556.5	8.5	384	2	09X602	09x602 streptomyc
28	554	8.5	471	3	09HEX8	09hex8 trichoderma
29	550	8.4	2232	5	P91365	P91365 caenorhabdi
30	542	8.3	996	2	09AQH0	09aqh0 caldicellul
31	540	8.3	457	3	093837	093837 acromonium
32	537	8.2	472	3	093860	093860 piromyces r
33	510.5	7.8	1915	2	09RPL0	09rpl0 acetivibrio
34	492.5	7.5	170	2	09RX6	09rx6 caldicellul
35	473.5	7.3	473	10	039620	039620 chlamydomon
36	452	6.9	458	3	09UW11	09uw11 piromyces r
37	444.5	6.8	449	3	P78721	P78721 orpiniomyces
38	443	6.8	460	3	09UW10	09uw10 piromyces r
39	436	6.7	428	3	012646	012646 neocallimas
40	431.5	6.6	997	2	09Z4I1	09z4i1 bacillus sp
41	431.5	6.6	2316	2	09FDJ9	09fdj9 bacteroides
42	419	6.4	260	2	033897	033897 rhodotherm
43	417	6.4	376	3	09P808	09p808 piromyces r
44	415.5	6.4	261	2	09AGC7	09agc7 caldicellul
45	412	6.3	459	3	P78720	P78720 orpiniomyces

ALIGNMENTS

RESULT ID	1	PRELIMINARY:	PRT:	683 AA.
09PDW2	09PDW2	09PDW2		
AC	01-OCT-2000 (Tremblrel. 15, Created)			
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)			
DE	1.4-BETA-CELLULOBIOSIDASE.			
GN	XF1267.			
OS	Xyella fastidiosa.			
OC	Bacteria: Proteobacteria; gamma subdivision; Xanthomonas group;			
OC	Xyella			
OX	NCBI_TaxID=2371;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=9A5C;			
RX	MEDLINE=20365717; PubMed=10910347;			
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,			
RA	Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,			
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britones M.R.S.,			
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carreiro D.M., Carter H.,			
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,			
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,			
RA	Feichman A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,			
RA	Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,			
RA	Gardner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,			
RA	Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,			
RA	Kriegler J.E., Kuramae E.E., Lalgeri F., Lambais M.R., Leite L.C.C.,			
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,			
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,			
RA	Martins M.V., Martins E.A.L., Martins E.M.F., Matsushima A.T.,			
RA	Menck C.F.M., Miracca E.C., Miyaki C.Y., Montello-Vitorello C.B.,			
RA	Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,			
RA	Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,			
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,			
RA	Peltozo B.R., Pereira G.A.G., Pereira H.A.Jr., Resguero J.B.,			
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,			
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,			
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,			

RA de Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Trufi D., Tsai S.M., Tsunako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Secubal J.C.,
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"
 RL Nature 406:151-159(2000).
 DR EMBL: AE003960; AAF84076.1; -
 DR HSP; P07987; ICB2.
 DR InterPro: IPR001524; Glyco_hydro_6.
 DR Pfam: PF01341; Glyco_hydro_6; 1.
 DR PRINTS: PR00733; GLHYDRLASE6.
 DR PRODOM: PD003733; Glyco_hydro_6; 2.
 KM Complete proteome.
 SQ SEQUENCE 683 AA; 70852 MM; DC8FB5F76B8E8D5F CRC64;

Query Match 20.8%; Score 1358; DB 16; Length 683;
 Best Local Similarity 43.9%; Pred. No. 2,6e-57;
 Matches 319; Conservative 81; Mismatches 218; Indels 108; Gaps 19;

QY 53 AATHVDNYAGATFFVNPYMAOEVOSEANQNTATLAKMRVSTYSTAAMDRITAIAG 112
 D 18 AEAHVDNPFVATSTYLNPDYSKIDTSLAKVNDVSLAKEMQTIKYPGVWLDITDAVKG 77
 QY 113 VNGGPG--LFTYDAALSOOGTPEVIEIYVDLPERDCAALASNGELPATAAGLQTY 169
 D 78 GPKNDGRNLGHDALAAOKNGKNTPATYATYDIPGRDCAALASNGELPTEGLDLY 137
 QY 170 ETQYIDPIASILSNPKYSLRIIVTIEPDSLPAVATNMSIOACATA--VPYEGGIEYAL 227
 D 138 KREYIDTASLFSNPKYKIDRIVNIIEPDSLPIITNMTSPSCAAKMTGIEBEIKAL 197
 QY 228 TKLAIIPVYIYMDAASGMLGMPNNAAGYVOEOKVL--NASIGVNGIDGFEVNTANTP 286
 D 198 NKLSIIPVYVYMDIGHGSLGMDNRIIPAVSLYKTKVIOSTAGPASYNGFATNTANTP 257
 QY 287 LKEPPM--PATOOGGAPVASEANFYQWNPDIADYADVLSYLAAGPSSIGMLIDTLR 345
 D 258 LIERLPPLDNLIGQPIRSSKFTYEMNRYPEMDYSEFLYNDYFAAGPSSIGFLIDTGR 317
 QY 346 NGMGGPNPTPTATDVNTFVNQSKIDLROHRLGIMCNGAGLQGPQASPTDPPNAHL 405
 D 318 NGMGGPERPTS--AFGNDVNSYVNSGRIDRRHNRGMCNQKAGIGLPIAT----PGGHV 372
 QY 406 DAYVIRKPPGSDGTSAASDPTTGKSDPCDPTTYSYGVLTNALPNSPIAGOWFPNAQF 465
 D 373 DAFQWIKPPGYSDGSSSLIPNDQKGFDRYCDPFTTTPDGYLTGALPAPLISGDFHQAQF 432
 QY 466 DQLVANARPAVPTSTSSPPPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 525
 D 433 VKLINNAIPDI-----SRTEPLPGSGSSSSSGSSSSSGSSSSSGSSSSSGSSSS 486
 QY 526 SPSSSP 585
 D 487 SGSSSSSSSSSGSSSSSGSSSSSGSSSSSGSSSSSGSSSSSGSSSSSGSSSSSS 546
 QY 586 GGLKQVYNNDSAPBDNLIKPLQLVNTGSSSVLDISTVYRWFTPDGSSSTLYVNCMA 645
 D 547 SG-----SSSSSGSSS-----SSGSSS-----GSSSGSGGN----- 572
 QY 646 AMGCG--NIRAFGVSYPATPATDYTLQLSFTGGTLAAGSGTEIGIONRKNKMSNFEEN 704
 D 573 -PGAGFNAGAF--NP-----GAGFNPGAGSDPTKNPIT-----SSITVDS 610
 QY 705 DYSYG-----TNTTFQDWTKVTVYNGVL--VMGT----- 732
 D 611 DMHTGTCERVKYNTNGSSRSRWSWT--VTIDPKGTIGTLMASATMSLGSGLIASGLDMNTL 669
 QY 733 ERSCTS 738
 D 670 EPNGT 675

RESULT 2
 ID 086730 PRELIMINARY; PRT; 579 AA.
 AC 086730;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PUTATIVE SECRETED CELLULASE.
 GN SC5C7.33.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Parkhill J., Barrett B.G., Rajandream M.A.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleiser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinasli H., Hopwood D.A.;
 RT "A 8 Mb Streptomyces coelicolor and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL051515; CAA20645.1; -
 DR HSP; P07986; IEXG.
 DR InterPro: IPR001919; CBD_2.
 DR InterPro: IPR001524; Glyco_hydro_6.
 DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF01341; Glyco_hydro_6; 1.
 DR PRINTS: PR00733; GLHYDRLASE6.
 DR PRODOM: PD003733; Glyco_hydro_6; 2.
 DR PROSITE: PS00655; GLYCOSYL_HYDROL_F6_1; 1.
 SQ SEQUENCE 579 AA; 60831 MM; 5A51DD28E7996A89 CRC64;

Query Match 16.1%; Score 1051; DB 2; Length 579;
 Best Local Similarity 49.2%; Pred. No. 9.5e-43;
 Matches 215; Conservative 61; Mismatches 131; Indels 30; Gaps 8;

QY 50 PAIATHVDNYAGATFFVNPYMAOEVOSEANQNTATLAKMRVSTYSTAAMDRITAI 109
 D 158 PTGPGDRVDNPDGQAVYVNPPEMSANAAEPGDR-----IADEPTGVWLDRIAA 207
 QY 110 INGVNGSGGLTYYDAALSOOGTPEVIEIYVDLPERDCAALASNGELPATAAGLQTY 169
 D 208 IEGANGGGLDHDAL--EQKSGSEMYYQVLYVLPDRDCSALASNGELGTE--IDRY 264
 QY 170 ETQYIDPIASILSNPKYSLRIIVTIEPDSLPAVATNMSIOACAT-----APYEGG 222
 D 265 KTEYIDPTAETISDSKVDLRIVTVEIDSLPNIVTNSGTPATENDVANKANGYOKG 324
 QY 223 IEYALTKHAIIPVYIYMDAASGMLGMPNNAAGYVOEOKVLNLSIG--NNGIDGFTN 280
 D 325 VGIALNRLGAVGNVYNYDAGHGLGWDN--FGPSAEIFYTAATTEGATLDVHGFTVN 383
 QY 281 TANTYPLKEPFWATQVGGQPVESANFYQWNPDIADYADVLSYLAAGPSSIGML 340
 D 384 TANTYALKEENKRTIDSVNGSVRSQSDVYDNNQYTDLSYQAMDKVLSIGFQONLGL 443
 QY 341 IDTLRNGWGPNPTPTATDVNTFVNQSKIDLROHRLGIMCNGAGLQGPQASPTDF 400
 D 444 IDTSRNGWGADRPYGPATPTDVNTYVNGGRYDRIRIHGNCNCGAGLGERPQASPA-- 501
 QY 401 PNAHLDAYVWIKPPGESDGTSAASDPTTGKSDPCDPTT---TSYGVLTNALPNSPIA 457

Db 502 --AGIDAYVWAKRPESDSSSKLIDNPDKGFRMCDPPTTGNERNGNSMGALPDAPIS 559
 458 GOMFPAOFDOLVANARP 474
 Db 560 GAMFSAQFOELMKNAFP 576

RESULT 3
 O9KH72 PRELIMINARY; PRT: 454 AA.
 ID O9KH72
 AC O9KH72: 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CELLULOSE PRECURSOR.
 GN EX.
 OS Thermomonospora fusca.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptosporangineae; Thermomonosporaceae;
 OC Thermobifida.
 OC NCBI_TaxID=2021;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Al Y.-C., Wilson D.B.;
 RT "Genomic DNA sequence encoding Thermomonospora fusca cellulase with
 overexpression."
 RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF268074; AAF75786.1; --
 DR HSSP; P07987; ICB2.
 DR Interpro; IPR001524; Glyco_hydro.6.
 DR Pfam; PF01341; Glyco_hydro.6; 1.
 DR PRINTS; PR00733; GLHYDRASE6.
 DR Prodom; PD003733; Glyco_hydro.6; 2.
 DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
 KW Signal.
 FT SIGNAL 1 31 POTENTIAL.
 FT CHAIN 32 454 CELLULOSE.
 SQ SEQUENCE 454 AA; 48624 MW; F8CD1C4CD4807134 CRC64;

Query Match 15.6%; Score 1015.5; DB 2; Length 454;
 Best Local Similarity 47.3%; Pred. No. 3.6e-41;
 Matches 221; Conservative 60; Mismatches 147; Indels 39; Gaps 12;

QY 27 LKAGVLAVGAVSTAASTVPLAMQHRAIAATH-----VDNPYAGAFEFNPPYMAQEVQSEAN 82
 DB 6 LRA--TLGAA--AALVSAALAPPSQAAAPGKVDNPEGAKLYNPVMSAKAAAPGCG 62
 QY 83 QTNATLAAMRYVSTYSTAVMMDRIAAING-----VNGGPGLTYYLDAALSOOGTTPREY 138
 DB 63 S-----AVANESTAVWMDRIGALIEGNSPTTSGMLDHLLEAV--ROSGGPRLTI 111
 QY 139 EIVYIDLPERDCAALASNGELPATAAGLOTYETQYIDPIASTLSN--PKYSSLRIVYIIEP 197
 DB 112 QVYVNLIPERDCAALASNGELGPDE--LDRIKSEYIDPIADIMDPADENLRIVAIIEI 169
 QY 198 DSLPNAVTKMSIQACATNPY-----YEGTEYALTKLHAIPNVYIYMDAAHSGMLGRP 251
 DB 170 DSLPNTLVTVNGSGTELCAYMKONGYVNGYALRKIGETIPNVYNYIDAHHGIGMD 229
 QY 252 NNASGYVEQVQVFLNAS--IGVNGIDGFTVNTANTYPLKEPFMTATQOVGQPVESANFYQ 310
 DB 230 SNFGPSVDIFYEAAANSSTGYVHGFTSNTANTYATPEYLDVNGTVNGQLRQSKWVD 289
 QY 311 WNPDDIDEAADVLSRLVAAEPSSIGMLDPLRNGMGKPNPEPSPATDVNTFVNOS 370
 DB 290 WQVYDELSTFYDRLQALLAKGFRSDIGMLIDTSRNGMGKPNRPTGSSSTDLNTIYVDES 349
 QY 371 KIDLROHRLGIMKONGAGIGOPPOASPTDFPNAHILDAVYVIRPGSGSDGTSAAADPTTGR 430
 DB 350 KIDRLRHGPNKMGQAGIGSERPYVPA-----PGVDATYVWAKRPESDSSSKLIDNPDKG 405

QY 431 KSDPMCDPPTTTSYGVLTN---ALPNSPIAGQFPAQFDOLVANARP 474
 Db 406 GFRMCDPPTTQGNARNGNPNPSGALPNAPISGHWFSAQFRELLANAYP 452

RESULT 4
 O60029 PRELIMINARY; PRT: 596 AA.
 ID O60029
 AC O60029:
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE BETA-1,4-EXOCELLULOSE PRECURSOR (EC 3.2.1.91).
 GN E3.
 OS Thermomonospora fusca.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptosporangineae; Thermomonosporaceae;
 OC Thermobifida.
 OC NCBI_TaxID=2021;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YX;
 RA Zhang S., Lao G., Wilson D.B.;
 RT "Characterization of a Thermomonospora fusca exocellulase."
 RL Biochemistry 34:0-0(1995).
 DR EMBL; U18978; AAA62211.1; --
 DR HSSP; P07986; IEXG.
 DR Interpro; IPR001919; CBD_2.
 DR Interpro; IPR001524; Glyco_hydro.6.
 DR Pfam; PF00553; CBD_2; 1.
 DR Pfam; PF01341; Glyco_hydro.6; 1.
 DR PRINTS; PR00733; GLHYDRASE6.
 DR Prodom; PD003733; Glyco_hydro.6; 1.
 DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
 KW Signal; Hydrolase; Glycosidase.
 FT SIGNAL 1 38 POTENTIAL.
 FT CHAIN 39 596 BETA-1,4-EXOCELLULOSE.
 SQ SEQUENCE 596 AA; 63547 MW; B0FA5277FE7721E0 CRC64;

Query Match 15.3%; Score 1000; DB 2; Length 596;
 Best Local Similarity 47.7%; Pred. No. 2.6e-40;
 Matches 210; Conservative 55; Mismatches 143; Indels 32; Gaps 9;

QY 50 PLIAATHVNPYAGAFEFNPPYMAQEVQSEANQTNATLAAMKRYVSTYAVMMDRIAA 109
 DB 172 PTNPGKVDNPEGAKLYNPVMSAKAAAPGGS-----AVANESTAVWMDRIG 221
 QY 110 ING-----VNGGPGLTYYLDAALSOOGTTPREYIYIDLPGHDCALASNGELPATAAG 165
 DB 222 IEGNSDPTTSGMLRDLHLEAV--ROSGGPRLTIQVYIYVNLPGHDCALASNGELGPDE-- 278
 QY 166 LOTYETQYIDPIASTLSN--PKYSSLRIVYIIEPDSLPAVNTKMSIQACATNPY----- 218
 DB 279 LDRYKSEYIDPIADIMDPADENLRIVAIIEIDSLPNLVTVNGSGTELCAYMKONG 338
 QY 219 YEGTEYALTKLHAIPNVYIYMDAAHSGMLGNPNNASGYVEQVQVFLNAS--IGVNGIDGP 277
 DB 339 YVNGVYALRKIGETIPNVYNYIDAHHGIGWDSNFPSPVDIFYEAAANSSTGYVHVGR 398
 QY 278 VNTANTYPLKEPFMTATQOVGQPVESANFYOMNPDDIDEAADVLSRLVAAEPSSI 337
 DB 399 ISNTAVYATPEYLDVNGTVNGQLRQSKWVDWQVDELSTFYDRLQALLAKGFRSDI 458
 QY 338 GMLIDPLRNGMGKPNPEPSPATDVNTFVNOSKIDLROHRLGIMKONGAGIGOPPOASP 397
 DB 459 GMLIDTSRNGMGKPNRPTGSSSTDLNTIYVDESIRDIRLHPGNMGQAGAGIGSERPTVP 518
 QY 398 TDEPNHILDAVYVIRPGSGSDGTSAAADPTTGRKSDPMCDPPTTTSYGVLTN---ALPNS 454
 DB 519 A-----PGVDATYVWAKRPESDSSSKLIDNPDKGFRMCDPPTTQGNARNGNPNPSGALPNA 574
 QY 455 PIAGQFPAQFDOLVANARP 474

RESULT	5	
AC	P96311	PRELIMINARY; PRT; 1711 AA.
AC	P96311	
DE	01-MAY-1997 (TREMBLrel. 03, Created)	
DE	01-MAY-1997 (TREMBLrel. 03, Last sequence update)	
DE	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	ENOGGLUCANASE A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE A)	
DE	(CELLULOSE A) (FRAGMENT).	
DE	CEL.	
OS	Anaerocellum thermophilum.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;	
OC	Anaerocellum group; Anaerocellum.	
OX	NCBI_TaxID=31899;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-Z-1320;	
RA	Zverlov V.;	
RL	Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.	
CC	-1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN	
CC	ENOGGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL	
CC	DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE	
CC	CELLULOSE (BY SIMILARITY).	
CC	-1- CATALYTIC ACTIVITY: ENDOPOLYMEROLYSIS OF 1,4-BETA-D-GLUCOSIDIC	
CC	LINKAGES IN CELLULOSE.	
CC	-1- SIMILARITY: THE N-TERMINAL PART BELONGS TO CELLULOSE FAMILY E	
CC	(FAMILY 9 OF GLYCOSYL HYDROLASES).	
CC	-1- SIMILARITY: THE C-TERMINAL PART BELONGS TO CELLULOSE FAMILY L	
CC	(FAMILY 48 OF GLYCOSYL HYDROLASES).	
DR	EMBL; Z86105; CAB06786.1; -.	
DR	HSSP; P26221; 1TF4.	
DR	InterPro: IPR001956; CBD_3.	
DR	InterPro: IPR000556; Glyco_hydro_48.	
DR	InterPro: IPR001701; Glyco_hydro_9.	
DR	Pfam: PF00942; CBD_3; 3.	
DR	Pfam: PF02011; Glyco_hydro_48; 1.	
DR	Pfam: PF00759; Glyco_hydro_9; 1.	
DR	PRINTS: PR00844; GLHYDRLASE48.	
DR	ProDom: PD001947; CBD_3; 2.	
DR	ProDom: PD011903; Glyco_hydro_48; 1.	
DR	PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.	
DR	PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.	
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.	
FT	NON_TER 1	
SEQUENCE	1711 AA; 189979 MW; E3E987CEB9CD0C21 CRC64;	
Query Match	11.4%; Score 743.5; DB 2; Length 1711;	
Best Local Similarity	22.8%; Pred. No. 1.4e-27;	
Matches 277; Conservative 169; Mismatches 353; Indels 417; Gaps		477;
QY	35 AVSIASIVPLAMQHPAIAATVNDVPYAGATFFVNPYMAQEVQSEA-ANQTNATLAAMKR	93
DB	156 AASLAASIVLKDNRPKATAYLOH-----AKELYERAEVTKSDAGTYAANG	202
QY	94 VYSTS-----TAVMMDRIAALINGVNGCGLTYYIDAALSQOQGTPEVEIEIYIDL	145
DB	203 YNSNSGTFDELSMAAVWL--YLATND-----STYLTKASIVYO-NMPKTSISNTIDY	252
QY	146 PGRCDAALASNGE--LPATMAGLQTYETQYIDPIAS-----ILSNPK-----Y	186
DB	253 KKAHQMDDVHNKAAALLKTKTKDKIYK-QIISHLDYWTGTYNGERIKTKTPGKLAMLDOW	311
QY	187 SLSRIVT-----IIEPDSLPAVNTNMSIOACATAVPPYEOGIEVAL-----TKLHA	232
DB	312 GSLRATATTAFLAFYSDPVGCPSTKKEIYR-----KFGESQIDVALGASGRSVYEGCT	366
QY	233 IENVYIYMDAAHSGM-----LGMPPNASSIVQDEVQKVLNASSIGVNCI	274

Db	367	NPBKRPHHRAHSSWMAOSQSPSYHRRHTLKGALVCGSGSDSYTTDDISNVYNNNEVADYN	426
QY	275	DGEYTTANTYPLKEPMTITQOYGGOVESANFYQNNPDIDEADYAVDLSYSLVAAEFP	334
Db	427	AGFEGALAK-----MTQLGCGNPIDFEKFAIE-TPTNDEF-----FVEGJIN	466
QY	335	SS-----IGMLIDTLANMGCGNREPIGSPSTADVNFVMSKIDLRQ--HRLGMCNQ--	384
Db	467	ASGTNFEIETAIYVN--OSGW-----PAKADKLKE--RYEVDISELIKAYSPQULT	515
QY	385	-----NGAGIGOPQASPTDFPNAHIDYVW-----IKPESDQTS	422
Db	516	LSTNYNGAKVSGP-----YVMAOSKRIYIIVDFGTGLTYLPGGDKYK	560
QY	423	ASDPTTKKSDPCMDPYTTTSYGVLYNALPNSPIAGQWFA--QFDOLVANARPAVPTST	481
Db	561	EVOEPIRAPQVNWQNDNSNDYSFODIKGVSSGSYVKTKYIPLYGDGVYVWMDGEGTSAT-	619
QY	482	SSPPPPPPSPASPSPPSPSPSSSPSSSPSSSPSPSPSPSPSPSPSPSPSPSPSPSP	541
Db	620	-----PPTATATPT	660
QY	542	SPSPSPSSSPSPSPSSSP	600
Db	661	TPPTPTATPTSTP-----TPSTSPSTPAAAGGQIKVLYANKETNST	700
QY	601	DNQIKPGLQVNTSSSVDLSVTYRYWMTFRDGGSSSTLYVNCDMAAGCCNIRASFGSVN	660
Db	701	TNTIRPMLKVVNTSSSIDLSRVTIRYMTYVDGKAQSAIS--DMAQIGASNVEKFKYLS	759
QY	661	PATPTATPTIQLSTGCT--LAAGSGTGEIQNRKNSKDSMNFETENDYST-GINTTPODM	717
Db	760	SSVSGADYILEIGKCSAGQIQAKQDKGEIQIRNKKDSMNTYMGNDMSWMOGMTYNGEN	819
QY	718	TKVYVYVNGVYVWMTSESGTSPSTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP	777
Db	820	VKVTAIYIDGVLYVWQGEBSGATPTTATPATPATVPTP-----	855
QY	778	LAMNASTDNGVAHYNVYRNGVLVGPVTYSFTDTGLAAGTAYTYVAADAAGNTSAPS	837
Db	856	-----TPTP-----S	861
QY	838	TPVATITTSPPSPSTPTPTGTYTDCPTPBNONGVTSVOGDERV-----QTNEMNSSAQOC	822
Db	862	TPPTATPATPTPTPTPTSTPVA-----GGQIKVLYANKETNSTTNTIRPW	906
QY	893	L-TINTFATGATVETANMSGGTGAPATPYSIYKCHGMCGTCKRNKMPQIQISQISAVT	951
Db	907	LKVYNTGSSSIDLSRVYIR-----YV-----YVGDKAQSAIS	940
QY	952	SWS-----TTOVSSGADVADADIWNTSPTTGGPNGTEIMI-----WLN	991
Db	941	DMAQIGASNVEKFKYLSVSGADYILEIGKCSAGQIQAKQDKTG-EIIRRNKSDMSN	999
QY	992	SRGVOPEGSOTATGVTYAGHTNANWOGQOTSW---KIISYVLTGPGATSI SNLDKAIF	104
Db	1000	YN-----QGNDS--WMQSMNTYENKVKVAYI-----	102
QY	1048	ADAAARGSLNTSDYLVEAGFELWOGGQIGLSFVSVTSQSSPTSPSPPTPTPTPT	110
Db	1026	-----DGVLVW--GOEBSGATPTTATPATPATVPTPTPTPTPTPTAT	106
QY	1108	PTSPSPPTSPSPPTSS 1123	
Db	1066	PTATPTPTPTPTPTSSPTS 1081	
RESULT	6		
Q9REF5	Q9REF5	PRELIMINARY;	PRT: 930 AA.
AC	Q9REF5;		
DT	01-MAY-2000 (TIREMBLrel. 13, Created)		
	01-MAY-2000 (TIREMBLrel. 13, Last sequence update)		

DT	01-DEC-2001	(TREMBLE, 19, Last annotation update)
DE	MULTIDOMAIN BETA-1,4-MANNANASE PRECURSOR.	
GN	MANA.	
OS	<i>Caldicellulosium cellulovorans</i> .	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;	
CC	<i>Bacillus/Staphylococcus</i> group; <i>Alcyclobacillus</i> group; <i>Caldicellulosium</i>	
OX	NCBI_TaxID=74586;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=20120520; PubMed=10653733;	
RA	Sunna A., Gibbs M.D., Chin C.W.T., Nelson P.J., Bergquist P.L.;	
RT	"A gene encoding a novel multidomain beta-1,4-mannanase from	
RT	<i>Caldicellulosium cellulovorans</i> and action of the recombinant enzyme on	
RT	kraft pulp.";	
RL	Appl. Environ. Microbiol. 66:664-670(2000).	
DR	EMBL: AF163837; AAF22274.1; --	
DR	HSSP; 006851; INBC.	
DR	InterPro: IPR001956; CBD_3.	
DR	InterPro: IPR004302; Chitin_bind_3.	
DR	InterPro: IPR001547; Glyco_hydro_F5.	
DR	InterPro: IPR002965; P_rich_extensn.	
DR	Pfam: PF00942; CBD_3; 2.	
DR	Pfam: PF00150; cellulase; 1.	
DR	Pfam: PF03067; Chitin_bind_3; 1.	
DR	PRINTS; PRO1217; PRICHEXTNSN.	
DR	ProDom: PD001947; CBD_3; 2.	
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.	
KW	Signal.	
FT	CHAIN 1 33	
FT	CHAIN 34 930	
FT	SEQUENCE 930 AA; 101576 MW; 008638D54DIAZCC CRC64;	
	POTENTIAL.	
	MULTIDOMAIN BETA-1,4-MANNANASE.	

Query Match	11.3%;	Score 737.5;	DB 2;	Length 930;
Best Local Similarity	25.0%;	Pred. No. 1.4e-27;		
Matches 246;	Conservative 111;	Mismatches 270;	Indels 357;	Gaps 36

```

0Y      411 IKPRESGTSASAPPTGRKSDMCDEITYT-----SYCVLINALPNPSIAOWE--- 461
Db      97 IIPBCKLCGPASPASTDGMNQARTD-----WMTTRQCPATITIVRYNAN--APHCEWTLLVY 149
0Y      462 -----PAQFOLVANMARPAVPTSTSSPPP-----PPSPSASPSPPSPSPSSSSPSP 509
Db      150 TRDWMDPPO-----PLMWSIDLEPRPFQOVYNPPIINSSGPCGAEXSMOVLQPNK 197
0Y      510 -----SPSP-----SSSPSPSPSPSPSSSPSPSSSPSPSPSPS 546
Db      198 OGRHIIYWIMORSDBSPFAFYKCSODYREGSGIAREFGDPREGCGTWITPPSPSGTTPPTPT 253
0Y      547 PPSPPSSPSPPSSSPSPSPSPSPSPSSPSDPTSPSPVSGGLKYOKXKNDSAPGDNOIKP 606
Db      258 PTPTSTPTP-----TPTPSVTPTPTPTSTPTSPS-----ASGTLREYRGVDTSATIDNMKP 308
0Y      607 GLIOLNTSSSVSDLSIVYRWTFMRDRDGSSLTYLNCCMAAGCNCINIRASFESVPAIPTA 666
Db      309 QLRINYISQAAPLTLELVKRYMITRN--STAEQYFCDDAQIGCSNIRQAFSLSPVSGA 367
0Y      667 DTYLOLSEGTGTLAAGSTGEIQNRVNKSIDSNETETNDYSY-CTNTFTPDWKRTVTVYN 725
Db      368 DSYIELSTGGGISIPAGCGNTEIQRIHRHTNNANNENETDNMSYNGAOQTWGPSTRILTYRN 427
0Y      726 GVLWVGCTPSTGSTSPRPSPPSPSPSPSGGVTTFRPVPTGLVTVGYSGSSVSLAMNSTD 785
Db      428 GVLWVGCTPSTGGSSRPPTVPTPTPTPTP----- 455
0Y      786 NVGVAAHYVNYENGVLVGOPTYTSETDTGLAAGTAYYTVAADVAAAGNTSAPSTEVTATT 845
Db      456 -----TS 457
0Y      846 SPSPSPPTGTYYTDCTEPCGNONGVTSVOGBDYRVQTNEM-----NSSAQCLINTATG 900
Db      458 TPTPTPTPTSTP-----TFSGGPNLSVNTQGVLGINHPHAMVRDLRLSSLDQ-----GIR 506

```

OY	AMVSTANESGCGATAPSPSYLKCGHMCNCTKKNVMDIOLSOIGSASMTSTVOSS	960
Db	507 SW-----GANAVALYLSNGCRW-----TK-----IPASEVADIIISARILGYRA	545
OY	961 GAYDV-----AYDIWTNSPTPTTGOPNGSTEIMILWLSNRGGVOPFGSOT	1003
Db	546 VLEVEHDTTGXGEDAALACSMTTAVNMYIELKNVLAOENFEYVNI-----GNEPFGNHN	599
OY	1004 ATG-----VYAGHTWNNWOGQOTSMKILISVILPGATSTISND-	1044
Db	600 YQNNVDTTRNAVALRNAGJNNTTIMVADPNN-----GQ--DW--SPTMKDNAPPTIFNADP	650
OY	1043 -LKAIFA-----DAAAR-----GSLNTS-----DYLLDY	1065
Db	651 QRNLVFSIHMVGYVDTAAEVOYSIESPVNGLPLVIGEGHMHSDDDPNEQAIVQYAOX	710
OY	1066 EAGETIMO-GGQIGJNSFSVSYSVTGTSPT-----PSPSPPTPPSPPTPPSP	1113
Db	711 NIGJFGSWSGNSGGVEYELDMVTNFNANSPTLAWGTFRINAIGTSTSPPTPTPTPTPT	770
OY	1113 SPSPSPSP-----TSSPSSGVCARATYVV-----	1137
Db	771 TPTPTPSAGCNLVYQYAAADTNATDNQKHFPIVNRGSSVPLSELITRTYWTYVDGDKP	830
OY	1138 ---NSDMGSGFATVTVNTGSRATNGMTVAWSFGNQTVTNMTALTALQSGASVATATNL	1194
Db	831 QVFMCDW-----AOVGSNLSRGSFVKLSTG--RTGADYVIEITFTSGAGSLAPCA	878
OY	1195 SYNNV-----IQPGSTTFGFGNGYS	1215
Db	879 SSGDIQVIRINKNDMTNTNEANDYS	902

RESULT	7		
09S3V4			
09S3V4	PRELIMINARY;	PRT:	351 AA.
AC	09S3V4;		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)		
DE	1.4-BETA-CELLULOBIOMYDROLASE A (EC 3.2.1.91) (FRAGMENT).		
CN	CEFA.		
OS	Cellulomonas flavigena.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas		
OX	NB01_TaxID1711;		
	[1]		
RM	SEQUENCE FROM N.A.		
RP	STRAIN-CDBB531;		
RA	Herrera A., Gutierrez A., Salgado L.M., Ponce-Novoa T.;		
RT	"Molecular characterization of cellulases from Cellulomonas		
RT	flavigena.";		
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF172344; AAD48493.1; -		
DR	HSSP; P07987; 1CB2.		
DR	InterPro; IPR001524; Glyco_hydro.6.		
DR	Pfam; PF01341; Glyco_hydro.6; 1.		
DR	PRINTS; PR00733; GLHYDRLAS6.		
DR	PRODOM; PD003733; Glyco_hydro.6; 1.		
DR	PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.		
FT	Hydrolase; Glycosidase.		
EW	NON_TER 351 351		
SO	SEQUENCE 351 AA; 37615 MW; CDEB63E5DF335C2 CRC64;		

Query Match	11.1%;	Score 722.5;	DB 2;	Length 351;
Best Local Similarity	48.1%;	Pred. No. 2.5e-27;		
Matches 166;	Conservative 45;	Mismatches 105;	Indels 29;	Gaps 13;

```
Oy      21 PAISRRLRACVLAGAVSIAASI--VPLAMQ--HPAIIA-THVDNRYAGATFEFVNPYWA-QE 75
|:::| | | : | : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      6 PSVARRRRTV--GAATTAAIVANPILTSLSPASAAEARVDNPYAGAVQVYNPNWASSS 63
```

[illegible]

RESULT	8	
09A0G4	PRELIMINARY;	PRT; 1751 AA.
ID 09A0G4		
AC 09A0G4;		
DT 01-JUN-2001	(TREMBLrel. 17, Created)	
DT 01-JUN-2001	(TREMBLrel. 17, Last sequence update)	
DT 01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE CELE.		
OS <i>Caldicellulosiruptor</i> sp. Tok7B.1.		
OC Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC Thermomicrobacter group; <i>Caldicellulosiruptor</i> .		
OX NCBI_Taxid=80339;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN-TOK7B.1;		
RX MEDLINE=20171169; Pubmed=10706665;		
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,		
RA Bergquist P.L.;		
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme		
RT thermophile <i>Caldicellulosiruptor</i> isolate Tok7B.1.";		
RL Curr. Microbiol. 40:333-340(2000).		
DR EMBL; AF078042; AAK06394.1; ..		
DR HSSP; 006851; INBC.		
DR InterPro; IPR001956; CBD_3.		
DR InterPro; IPR001701; Glyco_hydro_9.		
DR InterPro; IPR000566; Lipoclin_cyfabp.		
DR InterPro; IPR002965; P_tich_extensn.		
DR Pfam; PF00942; CBD_3; 1.		
DR Pfam; PF00759; Glyco_hydro_9; 1.		
DR PRINTS; PRO1217; PRICHEXTENS.		
DR ProDom; PD001947; CBD_3; 3.		
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.		
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.		
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.		
SO SEQUENCE 1751 AA; 192116 MW; 60178CBF3C00BE95 CRC64;		

	Score	DB 2	Length
Query Match	10.9%	708.5	1751
Best Local Similarity	22.3%	Pred. NO. 6.7e-26	
Best Local 304, Conservative	165	Mismatches 429	Indels 467
			Gaps 54

[illegible]

QY	151	-----AALASNGELPAPAAOLOTYEYOYIDPLASTILSNKYSLSLAVITIEPDISLP	201	
Db	366	INFSAQSIDVYLSTGRSFFVVGF- - -TNY-----POH-----P	396	
QY	202	NAVVTNMSIQOACATAVPYEEQIEVALTKLHAIRPVYIYMDAAHSGLGMPNNASGVOEY	261	
Db	397	HHRNAHSSMAWSMKIPEYHRHLTL-----GALVGSGSDSDINDI	437	
QY	262	OQVLNASIGVANGIDCFYTNTANTYPLKEPRMTATQQVGGOVPESANFYQMNPDIDEADYA	321	
Db	438	TDYVONEVACDYNAGIVGALKL-----MOLYGECEPIDDFKAIBE-TPTNDEI-FV	485	
QY	322	VDLYSRLVAAGEPSISGLIDLTELNGWGGNEPTGPSTADVTEFVNOSKIDILRQHGLM	381	
Db	486	ESEFNOSGPPMYETVISITY--RTGM-----PPTYDKLSF---KYFDLTDEL----	529	
QY	382	CNONAGLIGOPPOASPDPDFNAHLDA-----YVM-----IKPPGESD	418	
Db	530	-----IQAGSPDVVKVD--YYIEEGKISGPYVDMKNRNITYVLVDFSGTRKIYRGGEVE	582	
QY	419	GTSASDPDTGKKSKDDPKODPPTYTTSYGLVNLNALPNSEIAGOMFPPOFD--QLVANRRPAY	476	
Db	583	HKQAOEKTI SVPGCPW-DLPNDPSYGLISOLEKNR-----YIAVYDNMNLMWGLEPA	636	
QY	477	PSTSSTPPPPPSPASAPSPSPSPSSSPSPSPSPSSPSPSPSPSPSPSPSSPSPSPS	536	
Db	637	ATST-----PAT	644	
QY	537	SSPSPSPSPSPSPSSSPSPSPSSSPSPSPSPSPSSPSPSPSPSPSPSPSSPSPSPSVG-	GLKVQYKNN	595
Db	645	STPPTPTPTPTPTVATPPTP-----TPPTPTGSPGCGSVKLKYNK	686	
QY	596	DSAPEDNCKIRKGLOLVMTGSSSYDLSTVYRKRYFTBDGSGSLTVLYNCMDMAAGCGRAS	655	
Db	687	ELSASTGSIRMFPIKIVNGGSSVDLSKVKIRITYTDGDKPQSAV-CDMAQIGASNVPFN	745	
QY	656	FCSVNPATPADTYLQLSFTGTG--LAAGSTGEIONRVKNSDMSNFTEENDYS-GTWI	712	
Db	746	FUKLSSGVSAGADYLEVFESSGAGOLPQKDOTGDIQVRFKNMWSNMYNOADMWSLQSWT	805	
QY	713	TPMODTKTYVNVGVNWCMPEPSTSPSP-----PSPSPSPSPSPGCDVYP	760	
Db	806	NIGEAKATYLLVDGLVWGGPGGATPAPTSTATPTPTPTATPTPTPTPTPTPTPTPTPT	865	
QY	761	SVPTGLVYTVGSSSVSLANNASTDNGVAHVNYRNGVLVGPVTSTFDGLAAGTAY	820	
Db	866	PAPT--ASPYGGS-----YWPSE-----	882	
QY	821	TYTTVAAY--DAAGNTSABSTPTATTTSPSPSPPTGYTIDCTPCPNONGVTSVOGEDXR	879	
Db	883	STGALKVYVYANGNISLSP--NVLNPKRIKENVGTITAVDLS-----R	921	
QY	880	VOTNEM--NSSAOCCULINTATGA--WTYSTANFGSGTGAGATVPSTLYK--GCHMON	931	
Db	922	VKVRVYWIIDEAOTOSVVASINIPATIDYKFKLGANAGADYYIEIGFKSGAGVLAAG	981	
QY	932	CTTKNVGNPDISOIGSAVYTSWSTTOVSGAYVAVDIMTNSTPT-----TTG-----	979	
Db	982	OSTKRILRSI-----OKSGSXYNDSYSVRASNSYIENEEKVTGYIDVL	1026	
QY	980	-----OP-NGETIMIW-----LNSRGGOVOPPGSQTAGVYVAGHTYNNWOGOOT	1022	
Db	1027	WMGREPBNAOIKVYANGNLGSKNTYLNPKRIKENVGTJAVDSLHVKRVRYTTIDGEAT	1086	
QY	1023	SMKII SYVLJTFGATGISMLDKAIFADAARGLSLNTSDYLDV--BACEGIWOGO--	1076	
Db	1087	QSVSYTSSINP-----AYIDWKFKKLGANNAG--ADYVEIGEFSGAGVLAAGOSTIKE	1137	
QY	1077	-----GLG-----SNSESYSVTSG-----TSSPTSP	1096	
Db	1138	IRLSAQKSGSGSYNOSNBYSVANSAGYIENKRVYCIDGAIWVGREBSNRCKPAGCVTPTP	1197	

QY 1099 SPTPTSPPTPTSPPTSPPTSSPSSSGVACRAITYVNSDMGSGTATATVNTNGSRA 1158
 ID 096V97 PRELIMINARY; PRT; 490 AA.
 AC 096V97;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CELLOBIOHYDROLASE II-LIKE CELLULOSE CELI.
 GN CELI.
 OS Orpinomyces sp. PC-2.
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimastixales;
 OC Neocallimastixaceae; Orpinomyces.
 OX NCBI_TaxID=50059;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PC-2;
 RX MEDLINE=21405733; Pubmed=11514516;
 RA Steenbakkers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
 RA Ljungdahl L.G., Op Den Camp H.J.M.;
 RT "Noncatalytic docking domains of cellobiosomes of anaerobic fungi.";
 RL J. Bacteriol. 183:5325-5333(2001).
 DR EMBL: AF177205; AAL0121.1; -.
 KW Hydrolyase.
 SQ SEQUENCE 490 AA; 54051 MW; 85FA0D1280759886 CRC64;

RESULT 9
 ID 096V97 PRELIMINARY; PRT; 490 AA.
 AC 096V97;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CELLOBIOHYDROLASE II-LIKE CELLULOSE CELI.
 GN CELI.
 OS Orpinomyces sp. PC-2.
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimastixales;
 OC Neocallimastixaceae; Orpinomyces.
 OX NCBI_TaxID=50059;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PC-2;
 RX MEDLINE=21405733; Pubmed=11514516;
 RA Steenbakkers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
 RA Ljungdahl L.G., Op Den Camp H.J.M.;
 RT "Noncatalytic docking domains of cellobiosomes of anaerobic fungi.";
 RL J. Bacteriol. 183:5325-5333(2001).
 DR EMBL: AF177205; AAL0121.1; -.
 KW Hydrolyase.
 SQ SEQUENCE 490 AA; 54051 MW; 85FA0D1280759886 CRC64;

Query Match 10.8%; Score 705; DB 3; Length 490;
 Best Local Similarity 39.0%; Pred. No. 2, 4e-26;
 Matches 167; Conservative 54; Mismatches 147; Indels 60; Gaps 12;

QY 50 PAIATVNDNPYAGATFFNPYNAOEYSEANOTNATLAKKRVSTYSTAVMMDRIAA 109
 DB 118 PGEOYTHSGPFSGVEFFLNPYVAEYDAIEONTNSSLAKAKEMKTYNAIMLDYIK 176
 QY 110 INGVNGSGPLTTYDAALSOOGCTPEVIEI-VYIDLPGRCALASNGELPATAGLOT 168
 DB 177 ----NMHQMLETNKGALQOQEFEGSKVLVYFVYDLPGRDCHALASNGELANDSDAQR 232
 QY 169 YEQYIDPIASILSNPKYSLRYITIEPDSLPAVATNM-SIOACATVAPYEEGIEYAL 227
 DB 233 YKTEYIDVIEEKLK--YKSGQPVLLIIEPDSLANTVNTLSTPACRDSSEKYIYLDGHAAYLI 290
 QY 228 TKLHAIPNYIYMDAASHGLGMPNNASGYVOEKYVLNASIGVNG-IDGFEVNTANTYTP 286
 DB 291 OKLGVLPVAMVYLDIGHAFMLGMDNREKAKKYSKVI--SSGPGVYVKGTDVNAVYTP 348
 QY 287 LKPEFMTATQOVGQPVESANFYOMNPDIDEADYAVDLVSLVNAAGFPSSIGMLIDTLRN 346
 DB 349 WEDEPTLSRGFET-----EWNPCPEDEKRYLELMHKDFEAAKIESYEFYCDSRN 397
 QY 347 WGGPNEPTGPSTATDVNTVFNOSKIDLRHGRGLMCNONGAGLGQPPQASPTDFPNAHLD 406
 DB 398 G-----HKVD-RKHGEMCNOTGVGVARPOASPVSGMD-YLDA 433
 QY 407 AYWIKPGESEDGTSASADPTTGKSDPMDPPTTYSYGLVTLNALPNSPIAGOWEPAQFD 466
 DB 434 AFWIKRLGSDGTS-----DTSARIDGYCGH-----DTAKMKAPAEAGOMFOKHFE 480
 QY 467 QLVANARP 474
 DB 481 OGLENARP 488

Q96V98
 ID 096V98 PRELIMINARY; PRT; 491 AA.
 AC 096V98;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CELLOBIOHYDROLASE II-LIKE CELLULOSE CELI.
 GN CELI.
 OS Orpinomyces sp. PC-2.
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimastixales;
 OC Neocallimastixaceae; Orpinomyces.
 OX NCBI_TaxID=50059;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PC-2;
 RX MEDLINE=21405733; Pubmed=11514516;
 RA Steenbakkers P.J.M., Li X.L., Ximenes E.A., Arts J.G., Chen H.,
 RA Ljungdahl L.G., Op Den Camp H.J.M.;
 RT "Noncatalytic docking domains of cellobiosomes of anaerobic fungi.";
 RL J. Bacteriol. 183:5325-5333(2001).
 DR EMBL: AF177204; AAL0121.1; -.
 KW Hydrolyase.
 SQ SEQUENCE 491 AA; 53956 MW; 7C7ACC9BF946C4 CRC64;

Query Match 10.6%; Score 694; DB 3; Length 491;
 Best Local Similarity 38.6%; Pred. No. 8, 2e-26;
 Matches 165; Conservative 53; Mismatches 149; Indels 60; Gaps 12;

QY 50 PAIATVNDNPYAGATFFNPYNAOEYSEANOTNATLAKKRVSTYSTAVMMDRIAA 109
 DB 121 PGSOULTSGNPFSGVEFFLNPYVAEYDAIAQMSNSSLAKAKEMKTYNAIMLDYIK 179
 QY 110 INGVNGSGPLTTYDAALSOOGCTPEVIEIYIYDLPGRDCHALASNGELPATAGLOT 169
 DB 180 ----NMQOMLETNKGALQOQTSKVLVYFVYDLPGRDCHALASNGELANDSDAQR 235
 QY 170 ETOYIDPIASILSNPKYSLRYITIEPDSLPAVATNM-SIOACATVAPYEEGIEYALT 228
 DB 236 KTEYIDVIEEKLK--YKSGQPVLLIIEPDSLANTVNTLENTPACRDSQYIYLDGHAAYLI 293
 QY 229 KLAIPNYIYMDAASHGLGMPNNASGYVOEKYVLNASIGVNG-IDGFEVNTANTYTP 287
 DB 294 KFGVLPVAMVYLDIGHAFMLGMDNREKAKKYSKVI--SSGSPGKYRGFEDVNAVYTP 351
 QY 288 KEPMTATQOVGQPVESANFYOMNPDIDEADYAVDLXSLVNAAGFPSSIGMLIDTLRNG 347
 DB 352 EDPTLSRGFPT-----EWNPCPEDEKRYLEAHMKDFKAGI--SSYFVSDTSRNG 399
 QY 348 WGGPNEPTGPSTATDVNTVFNOSKIDLRHGRGLMCNONGAGLGQPPQASPTDFPNAHLD 407
 DB 400 -----HKTD-RKHGEMCNOTGVGIGARPPANPSSMD-YLDA 435
 QY 408 YWIKPGESEDGTSASADPTTGKSDPMDPPTTYSYGLVTLNALPNSPIAGOWEPAQFD 467
 DB 436 FYWIKRLGSDGTS-----DTSARIDGYCGH-----ETAKMKAPAEAGOMFOKHFEQ 482
 QY 468 LVANARP 474
 DB 483 GLENARP 489

RESULT 11
 ID 096V98 PRELIMINARY; PRT; 921 AA.
 AC 096V98;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE BETA-1,4-XYLANASE XYNA PRECURSOR.
 GN XYNA.
 OS Caldicellibacillus cellobiovarans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC	Bacillus/Staphylococcus group; Alicyclobacillus group; Caldicoccus
OX	NBRI_TaxID=74586;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=20519260; PubMed=11065373;
RT	Sunna A., Gibbs M.D., Bergquist P.L.;
RA	"A novel thermostable multidomain 1,4-beta-xylanase from
RT	'Caldicoccus celluloxans' and effect of its xylan-binding domain
RT	on enzyme activity";
RL	Microbiology 146:2947-2955(2000).
DR	EMBL; AF200304; AAF61649.1; "
DR	HSSP; Q06851; INBC.
DR	InterPro: IPR001956; CBD_3.
DR	InterPro: IPR001000; Glyco_hydro_10.
DR	InterPro: IPR002965; P_rich_extensn.
DR	Pfam; PF02018; CBD_6; 1.
DR	Pfam; PF00331; Glyco_hydro_10; 1.
DR	PRINTS: PR00134; G1HYDRLASE10.
DR	PRINTS: PR01217; PRICHEXTENS1.
DR	PRODOM: PD001947; CBD_3; 2.
KW	Signal; Xylan degradation; Hydrolase; Glycosidase.
FT	Signal
FT	CHAIN
SO	SEQUENCE 921 AA; 102380 MW; C5DDDA7E7567413 CR664;

Query Match	10.4%	Score 679;	DB 2;	Length 921;
Best Local Similarity	32.8%	Pred. No. 8.4e-25;		
Matches 165; Conservative	81;	Mismatches 107;	Indels 150;	Gaps 19;

OY	543	PSPBPSSSSPSPPSSSSPSPPSSPSPPSSPSPPSSPSPPSSPSGGKLVQYKKNDASAPDN	602
Dd	522	PSPPPTSTATPPTPTVLTPLTPPTPTPTSTPPTPTPS-ASGTLRVEKVDSSATDN	580
OY	603	QIRGGLDLYVTGSSSDVLSTVYARWFTBDGSGSTLYNCDMAAGCGNTRASFGSYNPA	662
Dd	581	QMRRQLRLRYNGSQAVALTELKARYWTKN-STOAEQYFCDMAOIGCSNRAOFVSILAOP	639
OY	663	TPTADTFLQLSFTGTGLAAGSGTEIGEIONRNKSDMSNFETENDSY-GTWTFPODMTKVT	721
Dd	640	VSGADSYIELSTFGTSVPAGGNGEIGNRHFTNNMNYNTEDMDSSNYCGLTTMGPSRIT	699
OY	722	VTVNGVLYMWGEBEG--TSBSPPTSPPSSPSBPGGVTPPYPTGLIVTVGVSGSYSLA	779
Dd	700	LVRNGVLYMWGEBEGGSSSTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT	744
OY	780	WNASTDWGAHVHYWRNGVLYVGOPYTSTFDGLAAGTAVTVAADAAGNTSASPSTP	839
Dd	745	-----PRTPT-----TSTP-TP	755
OY	840	VTATTTBSBPSPPTPTGTV-----TDCTGPBNQ-----NGVNS-----	872
Dd	756	TATPTPTPTPTPSAGMLVVOYRADPDNADDNLQPKPHFRIVNRGKTISVPLELTIKRWYT	815
OY	873	VQGGEYRQJNMENMSSAQCCLTINTATGAMTVSTANSGS-----TGCAEATY-----	920
Dd	816	VDGRKPQVFNCDM-----AM-VGCSNLRGSLVKLTITGRGCADYYLETFT	859
OY	921	---PSIYKGCHMGNCJTKRMGMPIQISOISIAVTSWSSTTVSSGAYVAADINTSNPTT	977
Dd	860	SGAGSLAPGANSCD-----IQARIK-----NDWTN-----YEANDYSID-----PTK	988
OY	978	TGQPNGTEIMYLNRS--GGYOP	998
Dd	899	TSTFADMNRVYTLYRNGOLVMGWEP	921

RESULT	12	
052374		
ID	052374	PRELIMINARY;
AC	052374;	PRT; 1779 AA.
DT	01-JUN-1998	(TREMBlrel, 06, Created)

DT 01-JUN-1998 (TMBBrel, 06, last sentence updated)
 DT 01-DEC-2001 (TMBBrel, 19, last annotation updated)
 DE FAMILY 10 XYLANASE (EC 3.2.1.8).
 GN XLYNC.
 OS *Caldicellulosiruptor* sp. RT69B.1.
 OC Bacteria; Firmicutes; Bacilli; Clostridium group;
 OC Thermocaneorobacter group; *Caldicellulosiruptor*.
 CX NCBI_taxonomy:70295;

RP SEQUENCE FROM N.A.
RC STRAIN=RT9B.1;
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RT "Family 10 and 11 xylanase genes from *Caldwellielloraptor* sp.
RT accession 1"

RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases
 DR EMBL: AF036924; AB95326.1; -.
 DR HSSP; Q06851; INBC.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR003305; CBD_6.
 DR InterPro: IPR001000; Glyco_hydro_10.
 DR Pfam: PF00942; CBD_3; 3.
 DR Pfam: PF02018; CBD_6; 2.
 DR Pfam: PF00331; Glyco_hydro_10; 1.
 DR PRINTS; PR00134; G1H7DRASE10.
 DR ProDom: PD001947; CBD_3; 3.
 DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
 KW Xylan degradation; Hydrolyase; Glycosidase.
 SQ SEQUENCE 1779 AA; 194304 MW; CE52698608B5CED CRC64

Query Match	10.3%	Score 675;	DB 2;	Length 1779;
Best Local Similarity	23.6%	Pred. NO. 2.7e-24;		
Matches 257;	Conservative 150;	Mismatches 416;	Indels 266;	Gaps 40;

```

OY      169 YEIDY-----IDPISILSNPYSSLSRLVITLIEBDSJLPNMTKMSIOACTANP--    217
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      659 YKSXTAWSLIETPVVPNSTLPAP-----AIQVPTSTPTPTPTPTPSATPTAPA7A    713
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

OY      218 -----YEEOGIEVALTKLHAIPNVYIYMDAAHSGMLGMPNNASGYOEOKVLMAISGV    271
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      714 SPAGSFWPSESSEYGALKV-----WY-----ANGMNSSTTNVLPKIKI    752
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

OY      272 NGIDGFYTNTANTPLKEPKNATIQOVGPSPESANFTQOMP-----DIDEAD    319
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      753 ENNG---TTAVDSLARKVKRWYTTIDGEAAOSVSAS--SINPAYIDRVVKLGANAAGAD    807
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

OY      320 YAVDLISRLVAAGFPSPSIGMLIDLTRNGMGNEPTGPSTADVTFNQSKIDLROHRG    379
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      808 YYEYV-----GFKSGAGVL-----AAGST-----KELRUSTOSSG    839
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

OY      380 LMCNONCAGLGGPPQAPSPTDFPNALH-DAYVIWKPGCESD-----GTSASDPYTKKS    432
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      840 SYNGSNDSYSRSANSYTIENKTYGVIDLVLMGRREPGNAQLKWYANGNLSSPTYNLPN    899
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

OY      433 DPMCDPYTTS-----YGVLTNALPNSPIAGCWFAQPD---OLIVANA-----    472
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      900 KIKIENGTTAAVDLSRKVKRYWTIDGEATQGSVASAINPAYIDVRVKLGANAGADY    959
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

OY      473 -----RPANPTISSPPPPPPSPSABS-----    486
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db      960 YVEYFGKSGAVLAGAQSTKEIRLSLOKSSGSGYNOSNDYSVRSANSYTIENEKYGYIDV    1019
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

OY      497 -----PSPPSPSSPSPPSPSSPSPPSPSPSSPSPPSSPSPPSSPSPPSPSPSPS    550
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      1020 LWMCKEPSKGTKPAAGEVTPAPPTSTPTPTPTPTPTTAPASAPPSB--TVRAATPTPTPTPT    1077
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

OY      551 SSPSPSSSSPSPPSPSPSSPSPPSSPSPPSPSVAG-SLKATYYKRNDSAPEDNOIKRGLO    609
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      1078 ATPPTPTT--TPTPTPTPTPTPTPTPTPTGPRGIGSLAKULYKNNETSASTGISIRWEK    1135
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

OY      610 LVNTGSSVDLSFVTVRYRWFTFRDGSSSTLYVNCDMAAMCGNIRAFSGSVNPATPTADTY    669
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      1136 IVNGSSSVSDLSVKKIRWYTVODGDGPQGAN-CDMAQIGASNVTNFENFKLLSGVSGADY    1194
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 16:16:30 : Search time 47.58 Seconds
(without alignments)
630,404 Million cell updates/sec

Title: US-09-917-384-1
Perfect score: 6525
Sequence: 1 MERTQSGRCRCRYRGTTTRM.....GFNGSYSGTNAAPLICTAS 1228

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	708.5	10.9	1751	4	US-09-136-574A-44
2	668.5	10.2	381	4	US-09-216-295-22
3	663.5	10.2	1426	4	US-09-136-574A-43
4	639.5	9.8	352	4	US-09-286-691-26
5	639.5	9.8	352	4	US-09-687-147-26
6	622.5	9.5	386	4	US-09-321-981-5
7	618	9.5	371	4	US-09-104-308-1
8	618	9.5	371	4	US-09-321-981-1
9	556.5	8.5	461	3	US-09-118-319-9
10	555	8.5	360	4	US-09-286-691-25
11	555	8.5	360	4	US-09-687-147-25
12	552	8.5	312	4	US-09-216-295-21
13	548.5	8.4	470	3	US-09-118-319-8
14	543	8.3	521	1	US-08-276-213-3
15	540	8.3	457	3	US-09-142-759-1
16	538	8.2	360	4	US-09-286-691-24
17	538	8.2	360	4	US-09-687-147-24
18	538	8.2	365	2	US-08-169-948B-12
19	538	8.2	365	2	US-08-448-873-12
20	538	8.2	365	4	US-08-382-452D-12
21	444.5	6.8	449	3	US-09-118-319-7
22	444.5	6.8	449	4	US-09-286-691-4
23	444.5	6.8	449	4	US-09-687-147-4
24	436	6.7	428	3	US-09-118-319-5
25	434.5	6.7	326	4	US-09-286-691-23
26	434.5	6.7	326	4	US-09-687-147-23
27	431.5	6.6	432	3	US-09-118-319-2

28	419	6.4	260	4	US-09-216-295-23	Sequence 23, Appl
29	417.5	6.4	551	2	US-09-033-537A-1	Sequence 1, Appl
30	413.5	6.3	616	4	US-09-136-574A-47	Sequence 47, Appl
31	412	6.3	459	3	US-09-118-319-6	Sequence 6, Appl
32	412	6.3	459	4	US-09-286-691-2	Sequence 2, Appl
33	412	6.3	459	4	US-09-687-147-2	Sequence 2, Appl
34	409	6.3	493	4	US-09-198-956-10	Sequence 10, Appl
35	409	6.3	493	4	US-09-198-955A-12	Sequence 12, Appl
36	381.5	5.8	167	5	PCT-US95-13813-9	Sequence 9, Appl
37	377.5	5.8	700	2	US-07-862-588B-2	Sequence 2, Appl
38	290	4.4	907	3	US-08-783-774-2	Sequence 2, Appl
39	280	4.4	907	5	PCT-US95-04611A-19	Sequence 19, Appl
40	284	4.4	348	4	US-09-216-295-16	Sequence 16, Appl
41	268.5	4.1	1185	4	US-09-041-886-23	Sequence 23, Appl
42	266	4.1	214	1	US-08-217-327-4	Sequence 4, Appl
43	259.5	4.0	259	4	US-09-216-295-5	Sequence 5, Appl
44	255.5	3.9	1719	2	US-08-459-568-4	Sequence 4, Appl
45	255.5	3.9	1719	2	US-08-399-411-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-136-574A-44
Sequence 44, Application US/09136574A
Patent No. 6294366
GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1751 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 44:

Db 342 QSGFGGTSSGAGFTAPACARLNGTSC 369

RESULT 9

US-09-118-319-9

Sequence 9, Application US/09118319

Patent No. 6114158

GENERAL INFORMATION:

APPLICANT: Li, Xin-Liang

APPLICANT: Chen, Huizhong

APPLICANT: Ljungdahl, Lars G.

TITLE OF INVENTION: Orpinomyces Cellulase Celf Protein and Coding Sequences

FILE REFERENCE: 33-98sequence listing

CURRENT APPLICATION NUMBER: US/09/118,319

CURRENT FILING DATE: 1998-07-17

NUMBER OF SEQ ID NOS: 9

SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 9

LENGTH: 461

TYPE: PRT

ORGANISM: Fusarium oxysporum

US-09-118-319-9

Query Match 8.5%; Score 556.5; DB 3; Length 461;

Best Local Similarity 32.3%; Pred. No. 7e-28; Indels 75; Gaps 16;

Matches 149; Conservative 71; Mismatches 167;

Db 17 TTRMPAISKRLRAGVLAVSIAISIVPLAQHPAIAATHVDNPYAGATFVNPYMAOEY 76

70 TAGPSTSTATKTATGSGSTTAGSVTSN---PPAAS---DNFYAGVDLMANNYTRSEV 123

77 QSEFA---NOTNATLAKMRVSTYSTAVWMRIAINGVNGSGPGLTLYDAALSOQOQT 133

124 MNLAVKLSGAKATAAKAVADVPSFO---WMDTYDHSILME---DTLADIRKANKAG- 174

134 TPVEIVIVYDLPGRCAALASNGELPATAAGTQYETQYIDIPASILSPKXSLRIYV 193

175 GKTAGPFTVYDLPNRDCAALASNGEYSLDKGANKRYKA-YIAIKIGLON--YSDRVIL 221

194 IIEPDSLPAVNTMSIOACATAVRYEGIEVALTKLHAIPNYIYMDAHSGLGMPNN 253

232 VIEPDSLANTVNTLNDKCAKASAKELVYAIKELN-LPNVSMILDAHGGLMGRAN 290

254 ASGYVEQVKVLNASIGVNGIDGFTVNTANTYPLKEPMTATQOVGQPVESANFYQWP 313

291 IGPAALYIAQIYDKAGKPSRVRGLVTNVSNN-----GWLKSTKPDYTESNP 337

314 DIDEADYAVDLYSRVLAAGPSSIGMLIDFLRMGSGPNEPTGPTATDVNTFVNQSKID 373

338 NYDEQRY-INAFAPLDAQEGSMVVKFIVDGRSG---KOPTG----- 375

374 LRORGLMCNONGAGLQOPQASPTDFPNAHLDAYVWIKRPGESDGTSAASDPTTGKSD 433

376 -QKAQGDWCNAKGTGFLRSTNTGD---ALADAFVWVKRGGESDGS-----DTSARKYD 427

434 PWCDPTTYSYGVJNALPNSPIAGQWFPAPQDQVLANARPA 475

428 YHCG-----LDALKPAPAGTWFQAYFQQLDNNAPS 459

Db 375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

375

CURRENT FILING DATE: 1999-04-05

EARLIER APPLICATION NUMBER: US 60/027,883

EARLIER FILING DATE: 1996-10-04

EARLIER APPLICATION NUMBER: PCT US97/18008

EARLIER FILING DATE: 1997-10-03

NUMBER OF SEQ ID NOS: 29

SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 25

LENGTH: 360

TYPE: PRT

ORGANISM: Fusarium oxysporum

US-09-286-691-25

Query Match 8.5%; Score 555; DB 4; Length 360;

Best Local Similarity 33.7%; Pred. No. 6.3e-28;

Matches 142; Conservative 64; Mismatches 147; Indels 68; Gaps 14;

Db 58 DNPYAGATFVNPYMAOEYSEAA---NOTNATLAKMRVSTYSTAVWMRIAINGV 114

3 DNPYAGVDLMANNYTRSEVNNLAVPKLSGAKATAAKAVADVPSFO---WMDTYDHSILME 59

115 GCGGLTLYDAALSOQOQTTPVEIVIVYDLPGRDCAALASNGELPATAAGTQYETQYI 174

60 -----DTLADIRKANKAG-GKTAGQFVYDLPNRDCAALASNGEYSLDKGANKRYA-YI 112

175 DPASILSNPKYSLSRIWTIIEPDSLPAVNTMSIOACATAVRYEGIEVALTKLHAIP 234

113 AKIKIGLON--YSDTFVYIIVIEPDSLANTVNTLNDKCAKASAYKELVYAIKELN-LP 169

235 NYIYMDAHSGLGMPNNASGYVEQVKLNASIGVNGIDGFTVNTANTYPLKEPMTA 294

170 NVSMYLDAGHGWLGMWANGPAKLYIAQIYDKAGKPSRVRGLVTNVSNN----- 220

295 TQOVGQPVESANFYQWNPDIIDEADYAVDLYSRVLAAGFPSSIGMLIDFLRMGSGPNEP 354

221 -----GWLKSTKPDYTESNPYDEQRY-INAFAPLDAQEGSMVVKFIVDGRSG-----KOP 271

355 TGPSSTATDVNTFVNQSKIDLRORGLMCNONGAGLQOPQASPTDFPNAHLDAYVWIKPP 414

272 TG-----QKAQGDWCNAKGTGFLRSTNTGD---ALADAFVWVKPG 310

415 GESDGTSAASDPTTGKSDMCPPTTTSYGVJNALPNSPIAGQWFPAPQDQVLANARP 474

311 GESDGT-----DTSARKYDHCG-----LDALKPAPAGTWFQAYFQQLDNNANP 357

475 A 475

358 S 358

Db 358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358

358


```

Oy      434  PMCDPTTSTYGLTNALPNSPIAGQWEPQADOLVANARPAV 476
         |      ::|| ::|| ::|| ::|| ::|| ::|| ::|| ::||
Db      423  FHCG-----YSDALQAPAEAGTWEQAVFYQLLTNNANPAL 456

```

Search completed: August 29, 2002, 16:22:32
Job time: 362 sec

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)